U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-255 301

Growth Effects of Major Land Use Projects: Volume I
Spec. and Causal Analysis of Model

Abcor, Inc.

Prepared For
Environmental Protection Agency

May 1976

KEEP UP TO DATE

Between the time you ordered this report—which is only one of the hundreds of thousands in the NTIS information collection available to you—and the time you are reading this message, several *new* reports relevant to your interests probably have entered the collection.

Subscribe to the **Weekly Government Abstracts** series that will bring you summaries of new reports *as soon as they are received by NTIS* from the originators of the research. The WGA's are an NTIS weekly newsletter service covering the most recent research findings in 25 areas of industrial, technological, and sociological interest—invaluable information for executives and professionals who must keep up to date.

The executive and professional information service provided by NTIS in the **Weekly Government Abstracts** newsletters will give you thorough and comprehensive coverage of government-conducted or sponsored research activities. And you'll get this important information within two weeks of the time it's released by originating agencies.

WGA newsletters are computer produced and electronically photocomposed to slash the time gap between the release of a report and its availability. You can learn about technical innovations immediately—and use them in the most meaningful and productive ways possible for your organization. Please request NTIS-PR-205/PCW for more information.

The weekly newsletter series will keep you current. But *learn what you have missed in the past* by ordering a computer **NTISearch** of all the research reports in your area of interest, dating as far back as 1964, if you wish. Please request NTIS-PR-186/PCN for more information.

WRITE: Managing Editor
5285 Port Royal Road
Springfield, VA 22161

Keep Up To Date With SRIM

SRIM (Selected Research in Microfiche) provides you with regular, automatic distribution of the complete texts of NTIS research reports *only* in the subject areas you select. SRIM covers almost all Government research reports by subject area and/or the originating Federal or local government agency. You may subscribe by any category or subcategory of our WGA **(Weekly Government Abstracts)** or **Government Reports Announcements and Index** categories, or to the reports issued by a particular agency such as the Department of Defense, Federal Energy Administration, or Environmental Protection Agency. Other options that will give you greater selectivity are available on request.

The cost of SRIM service is only 45¢ domestic (60¢ foreign) for each complete microfiched report. Your SRIM service begins as soon as your order is received and processed and you will receive biweekly shipments thereafter. If you wish, your service will be backdated to furnish you microfiche of reports issued earlier.

Because of contractual arrangements with several Special Technology Groups, not all NTIS reports are distributed in the SRIM program. You will receive a notice in your microfiche shipments identifying the exceptionally priced reports not available through SRIM.

A deposit account with NTIS is required before this service can be initiated. If you have specific questions concerning this service, please call (703) 451-1558, or write NTIS, attention SRIM Product Manager.

This information product distributed by

NTIS U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE Growth Effects of Major Land Use Projects. Vol. I: Specification and Causal Analysis of Model		5. REPORT DATE May, 1976 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Frank Benesh, Peter Guldberg, Ralph D'Agostino		8. PERFORMING ORGANIZATION REPORT NO. C-781 - a
9. PERFORMING ORGANIZATION NAME AND ADDRESS Walden Research Division of Abcor 201 Vassar Street Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO. 68-02-2076
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Protection Agency Office of Air Quality Planning and Standards Strategies and Air Standards Division (MD-12) Research Triangle Park, North Carolina 27711		13. TYPE OF REPORT AND PERIOD COVERED Final 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Growth Effects of Major Land Use Projects is a research program whose goal is to formulate a methodology to predict air pollutant emissions resulting from the construction and operation of two types of major land use projects, large residential projects and large concentrations of employment (i.e., office parks and industrial parks). Emissions are quantified from the major project, from land use induced by the major project, from secondary activity occurring off-site (i.e., generation of electricity by utilities), and from motor vehicle traffic associated with both the major project and its induced land uses.

This report documents the development of a model to predict the induced land use from such major projects. The report discusses the theoretical basis of the model, the specification of the equations that model the theory, and the causal analysis of the equations. Sample selection and data collection is also discussed.

Subsequent reports (i.e., Volume II and Volume III) document the emission factors, the land use model predictive equations and the traffic model.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Land Use Planning Industrial Areas Residential Areas	Path Analysis Causal Analysis Secondary Effects Induced Land Use	

PRICES SUBJECT TO CHANGE

18. DISTRIBUTION STATEMENT Unlimited	19. SECURITY CLASS *(This Report)* Unclassified	21. NO. OF PAGES 222
1.	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE 7.75-2.25

EPA Form 2220-1 (9-73)

212148

PB 255 301

EPA-450/3-76-012-a

May 1976

GROWTH EFFECTS OF MAJOR LAND USE PROJECTS: VOLUME I - SPECIFICATION AND CAUSAL ANALYSIS OF MODEL

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Waste Management
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

EPA-450/3-76-012-a

GROWTH EFFECTS OF MAJOR LAND USE PROJECTS: VOLUME I - SPECIFICATION AND CAUSAL ANALYSIS OF MODEL

by

Frank Benesh, Peter Guldberg, and Ralph D'Agostino

Walden Research Division of Abcor
201 Vassar Street
Cambridge, Massachusetts 02139

Contract No. 68-02-2076

EPA Project Officer: Thomas McCurdy

Prepared for

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air and Waste Management
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

May 1976

This report is issued by the Environmental Protection Agency to report technical data of interest to a limited number of readers. Copies are available free of charge to Federal employees, current contractors and grantees, and nonprofit organizations - as supplies permit - from the Air Pollution Technical Information Center, Environmental Protection Agency, Research Triangle Park, North Carolina 27711; or, for a fee, from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.

This report was furnished to the Environmental Protection Agency by Walden Research Division of Abcor, Cambridge, Massachusetts 02139, in fulfillment of Contract No. 68-02-2076. The contents of this report are reproduced herein as received from Walden Research Division of Abcor. The opinions, findings, and conclusions expressed are those of the author and not necessarily those of the Environmental Protection Agency. Mention of company or product names is not to be considered as an endorsement by the Environmental Protection Agency.

Publication No. EPA-450/3-76-012-a

ACKNOWLEDGEMENTS

Special appreciation goes to Mr. Thomas McCurdy, the U.S. Environmental Protection Agency project officer for this study, whose extensive assistance and advice was indespensible in the execution of this study.

In addition, we wish to express our appreciation to the cooperation of the more than one hundred individuals who were contacted and provided information during the data collection phase of this project.

STAFFING

Mr. Frank Benesh was the project manager of this study at Walden Research. Dr. Ralph D'Agostino, contributed to the overall study design and analysis. Mr. Mahesh Shah assisted in data collection and Mr. Peter Guldberg conducted much of the causal analysis. The data collection was coordinated by Mrs. Allison B. Goodsell and the manuscript was prepared under the direction of Ms. Gail Kelleher.

Metcalf and Eddy of Boston were subcontractors to Walden Research, assisting in model specification, sample selection, and data collection. Mr. Richard Ball, Mrs. Elizabeth Levin, Mr. Stephen Koop, Mrs. Nancy Lundgren, and Mr. Gerald Takano contributed to Metcalf and Eddy's effort.

TABLE OF CONTENTS

Chapter Title Page

I INTRODUCTION, APPROACH AND SUMMARY OF RESULTS 1-1

- A. Introduction 1-1
- B. Organization of the Report 1-2
- C. Approach 1-3
  - 1. Theory of Induced Development 1-3
  - 2. Selection of General Approach 1-3
  - 3. Statement of Fundamental Model 1-5

II SUMMARY OF STATISTICAL TECHNIQUES (PATH ANALYSIS) 2-1

- A. Aim of Path Analysis 2-1
- B. Path Diagrams 2-1
- C. Approach (Structural Model). 2-3
- D. Connection with Regression 2-4
- E. Assumptions and Deviations 2-4
  - 1. Linearity. 2-4
  - 2. Recursive 2-5
  - 3. Causal Priorities 2-7
  - 4. Uncorrelated Residuals of Endogenous Variables . 2-8
  - 5. Stability of the Estimates 2-9
  - 6. Multicollinearity not Present 2-10
  - 7. Scale 2-11
  - 8. Suppressor Variable Problem 2-11
- F. Other Assumptions 2-12
- G. Theory Trimming 2-14

III INITIAL LAND USE MODEL SPECIFICATION

- A. Objective 3-1
- B. Definition of Major Project 3-1
- C. Area of Influence 3-2
  - 1. Should the area of influence vary with the size or density of the major project?. 3-2
  - 2. What is the appropriate size of the study area (area of influence) for each major project size? 3-5
- D. Identification 3-16
- E. Methodology 3-19
- F. Model Description 3-21
  - 1. Endogenous Variables 3-21
  - 2. Exogenous Variables 3-28
  - 3. Simultaneous Block 3-31
  - 4. Recursive Block 3-31
- G. Theoretical and Empirical Basis for the Model . . 3-31
  - 1. Equation 1 3-31
  - 2. Equation 2 3-44

TABLE OF CONTENTS (CONTINUED)

Chapter | Title | Page

3. Equation 3 3-46
4. Equation 4 3-49
5. Equation 5 3-51
6. Equation 6 3-52
7. Equation 7 3-53
8. Equation 8 3-55
9. Equation 9 3-55
10. Equation 10 3-56
11. Equation 11 3-57
12. Equation 12 3-58

IV SAMPLE SELECTION 4-1

A. Purpose and Initial Criteria. 4-1
B. Methodology 4-1
C. Revised Criteria 4-2
D. Selection Process 4-3

V DATA COLLECTION 5-1

A. White Westinghouse and Western Electric, Columbus, Ohio 5-1
B. Montclaire - Starmont, Charlotte, North Carolina . . 5-25
1. Photograph Scale 5-25
2. Data Aggregation 5-25

VI CAUSAL ANALYSIS 6-1

A. General Approach 6-1
B. Data Transformations 6-4
C. Problems in Path Analysis 6-16
1. Multicollinearity 6-16
2. Suppressor Variables 6-19
3. Multiple Variable Definitions 6-22
4. Instrumental Variables 6-24
D. Path Diagrams 6-25
E. Theory Trimming 6-35
1. Approach 6-35
2. Discussion of Individual Equations 6-37
F. Comparison of Model Forms 6-60
G. Stability of Path Coefficients 6-63
1. Extreme Value Index 6-65
2. Coefficient Variation Index 6-74
H. Conclusions 6-77

VII REFERENCES 7-1

VIII APPENDICES

A. Bibliography of Land Use Models and Land Use Development A-1
B. Plots of Regresssion Coefficients for the Twenty Jackknifing Runs B-1

LIST OF TABLES

Table	Title	Page
3-1	INITIAL LIST OF ENDOGENOUS VARIABLES	3-22
3-2	MODEL VARIABLES AND DEFINITIONS	3-25
3-3	FIVE EQUATION SIMULTENEOUS BLOCK IN THE RESIDENTIAL MODEL	3-32
3-4	FIVE EQUATION SIMULTANEOUS BLOCK IN THE INDUSTRIAL-OFFICE MODEL	3-32
3-5	SEVEN EQUATION RECURSIVE BLOCK IN THE RESIDENTIAL MODEL	3-35
3-6	SEVEN EQUATION RECURSIVE BLOCK IN THE INDUSTRIAL-OFFICE MODEL	3-35
3-7	ORIGINAL SPECIFICATION OF THE RESIDENTIAL MODEL	3-38
3-8	ORIGINAL SPECIFICATION OF THE INDUSTRIAL-OFFICE MODEL. .	3-39
5-1	LIST OF DATA COLLECTION ITEMS	5-2
5-2	WESTERN ELECTRIC/WHITE WESTINGHOUSE	5-24
5-3	COMPARISON OF BUILDING AREA DATA (10^3 SQ.FT.)	5-26
5-4	CUMULATIVE PERCENTAGE OF CASE STUDY DATA COLLECTION OBTAINED FROM PHOTOGRAPHS OF CERTAIN SCALE OR LARGER. .	5-27
6-1	PATH ANALYSIS MODEL DATA TRANSFORMATIONS	6-5
6-2	RESIDENTIAL MODEL SIMPLE CORRELATIONS SIGNIFICANT AT $p < .10$	6-17
6-3	INDUSTRIAL-OFFICE MODEL SIMPLE CORRELATIONS SIGNIFICANT AT $p < .10$	6-18
6-4	ANALYSIS OF SIMPLE CORRELATIONS BETWEEN VARIABLES IN THE RESIDENTIAL PATH MODEL	6-20
6-5	ANALYSIS OF SIMPLE CORRELATIONS BETWEEN VARIABLES IN THE INDUSTRIAL-OFFICE MODEL	6-21
6-6	INSTRUMENTAL VARIABLES USED IN THE FIRST PATH ANALYSIS.	6-26
6-7	REDEFINITION OF PATH ANALYSIS MODEL DATA TRANSFORMATIONS	6-38
6-8	GROSS PRIVATE DOMESTIC INVESTMENT IN THE UNITED STATES, 1955-1973, IN BILLIONS OF 1958 DOLLARS	6-50
6-9	COMPARISON OF PATH REGRESSION COEFFICIENTS FROM THE FINAL PATH ANALYSIS AND THE JACKKNIFING STABILITY ANALYSIS FOR THE SIMULTANEOUS BLOCK OF EQUATIONS	6-66
6-10	EXTREME VALUE INDEX OF STABILITY FOR ALL PATH REGRESSION COEFFICIENTS IN THE SIMULTANEOUS BLOCKS OF EQUATIONS .	6-71
6-11	STABILITY INDEX FOR ALL PATH REGRESSION COEFFICIENTS IN THE SIMULTANEOUS BLOCKS OF EQUATIONS BASED ON THE EFFECT OF VARIATION IN THE COEFFICIENTS ON THE DEPENDENT VARIABLE	6-76

LIST OF FIGURES

Figure	Title	Page
2-1	ILLUSTRATIVE PATH DIAGRAMS	2-2
2-2	LEAD - LAG RELATIONSHIPS	2-6
2-3	ILLUSTRATION OF SUPPRESSOR VARIABLE	2-13
3-1	ILLUSTRATION OF EFFECT OF DIFFERING MAJOR PROJECT SIZES	3-3
3-2	RELATIONSHIPS OF LAND USE TO DISTANCE FROM THE AIRPORT	3-7
3-3	ZONES USED FOR LAND USE/DISTANCE ANALYSIS	3-8
3-4	FIVE EQUATION SIMULTANEOUS BLOCK IN THE RESIDENTIAL MODEL	3-33
3-5	FIVE EQUATION SIMULTANEOUS BLOCK IN THE INDUSTRIAL-OFFICE MODEL	3-34
3-6	SEVEN EQUATION RECURSIVE BLOCK IN THE RESIDENTIAL MODEL	3-36
3-7	SEVEN EQUATION RECURSIVE BLOCK IN THE INDUSTRIAL-OFFICE MODEL	3-37
3-8	ORIGINAL SPECIFICATION OF THE RESIDENTIAL MODEL . . .	3-40
3-9	ORIGINAL SPECIFICATION OF THE INDUSTRIAL-OFFICE MODEL	3-41
5-1	WESTERN ELECTRIC CASE STUDY AREA AND MORPC TRANSPORTATION ZONES	5-22
5-2	WHITE-WESTINGHOUSE CASE STUDY AREA AND MORPC TRANSPORTATION ZONES	5-23
6-1	ORIGINAL RESIDENTIAL PATH DIAGRAM	6-27
6-2	ORIGINAL INDUSTRIAL-OFFICE PATH DIAGRAM	6-28
6-3	INITIAL RESIDENTIAL PATH DIAGRAM	6-29
6-4	INITIAL INDUSTRIAL-OFFICE PATH DIAGRAM	6-30
6-5	INITIAL PATH ANALYSIS FOR THE RESIDENTIAL MODEL . .	6-31
6-6	INITIAL PATH ANALYSIS FOR THE INDUSTRIAL-OFFICE MODEL	6-32
6-7	FINAL PATH ANALYSIS FOR THE RESIDENTIAL MODEL . . .	6-33
6-8	FINAL PATH ANALYSIS FOR THE INDUSTRIAL-OFFICE MODEL	6-34
6-9	PATH ANALYSIS FOR THE RESIDENTIAL MODEL IN MULTIPLICATIVE FORM	6-61
6-10	PATH ANALYSIS FOR THE INDUSTRIAL-OFFICE MODEL IN MULTIPLICATIVE FORM	6-62
6-11	HYPOTHETICAL LINEAR REGRESSION	6-73

LIST OF FIGURES (CONTINUED)

Figure	Title	Page
B-1	PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE RES EQUATION IN THE RESIDENTIAL MODEL	B-2
B-2	PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE COMM EQUATION IN THE RESIDENTIAL MODEL	B-3
B-3	PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE OFFICE EQUATION IN THE RESIDENTIAL MODEL	B-4
B-4	PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE MANF EQUATION IN THE RESIDENTIAL MODEL	B-5
B-5	PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE HWLMNX EQUATION IN THE RESIDENTIAL MODEL	B-6
B-6	PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE RES EQUATION IN THE INDUSTRIAL/OFFICE MODEL	B-7
B-7	PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE COMM EQUATION IN THE INDUSTRIAL/OFFICE MODEL	B-8
B-8	PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE OFFICE EQUATION IN THE INDUSTRIAL/OFFICE MODEL	B-9
B-9	PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE MANF EQUATION IN THE INDUSTRIAL/OFFICE MODEL	B-10
B-10	PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE HWLMNX EQUATION IN THE INDUSTRIAL/OFFICE MODEL	B-11

LIST OF EXHIBITS

Exhibit	Title	Page
3-1	INDUCED RESIDENTIAL EXAMPLE	3-11
3-2	INDUCED COMMERCIAL DEVELOPMENT	3-13
4-1	EPA LETTER TO REGIONAL PLANNING AGENCIES	4-5
5-1A	ON SITE VISIT DATA COLLECTION FORMS	5-8
5-1B	CENSUS OF POLLUTION AND HOUSING DATA COLLECTION FORMS	5-15

I. INTRODUCTION, APPROACH AND SUMMARY OF RESULTS

A. INTRODUCTION

This report documents the results of the first three phases of a study of the growth effects of major land use projects (GEMLUP) undertaken by Walden Research. The principal objectives of the GEMLUP study are to formulate a methodology to predict air pollutant emissions from:

- two types of major land use developments: large concentrations of employment such as office or industrial parks, and large residential developments
- secondary land development that is induced by the two types of major land use development projects
- motor vehicular traffic associated with both the major project and secondary development.

GEMLUP relates to a number of EPA programs, including air quality maintenance plan (AQMP) development [1], environmental impact statement (EIS) review [2], the indefinitely suspended portions of indirect source review [3], and the prevention of significant air quality deterioration, or non-degradation [4]. Explicit or implicit in these programs in an evaluation of air quality impacts of land use plans or project developments. GEMLUP is designed to formulate and test a method of evaluating land use impacts at the project scale, and, in the process, develop a set of land use based emission factors potentially useful at the regional scale.

The study was divided into six phases

Phase 1 - specification of a preliminary model and generation of a list of data requirements

Phase 2 - data collection

Phase 3 - causal analysis of the land use model using path analysis

Phase 4 - development of predictive equations for the land use model and development of a traffic model

Phase 5 - development of indices of fuel consumption

Phase 6 - translation of fuel consumption indices into land use based emission factors.

The ability to accurately predict the secondary development of a major land use development alone is insufficient; it is also necessary to understand the complex interrelationships inherent in a predictive model. Therefore, after the formulation of a theory of secondary development, two distinct phases of analysis followed: Phase 3) the testing of the theory using path analysis, and Phase 4) the calibration of the theoretical relationships as predictive equations. This report concerns itself with the former analysis. The latter analysis (Phase 4) will be included in volume III of the final report.

In addition, a separate technical report (Volume II of the final report) will be prepared documenting the land use based emission factors (Phase 5 and 6). Volume III of the final report will document the application of the land use model, traffic model, and land use based emission factors in predicting air pollutant emissions from major development projects and their secondary effects.

Two additional appendices to this report (Volume I) has been prepared separately [42]. Appendix C documents the land use data and simple (zero order) correlations between that data. Appendix D presents the statistical output of the final path analytic model.

B. ORGANIZATION OF THE REPORT

The remainder of this introductory chapter discusses the general approach of the study and briefly summarizes the findings. Following this chapter, the statistical techniques employed in this study are reviewed in Chapter II. Chapter III contains the specification of the model, i.e., the translation of the general theory outlined above into explicit relationships. The next two chapters, IV and V, summarize respectively, the sample selection and data collection processes. Chapter VI contains the results of the path analysis, i.e., the testing of the causal relationships explicit in the model and discusses the results of the causal analysis.

C. APPROACH

1. Theory of Induced Development

Taking the industrial/office major land use project type as the more general case of the two types investigated, we adopted the following theory of induced development.*

Constructing a large source of employment like an industrial/office complex generates jobs which result in the nearby construction of dwelling units; these induce retail development to locate near them and generate demand for community, cultural, and religious facilities (schools, recreation areas, libraries, churches, theaters, fire and police stations, etc.). All of this requires the construction of streets and highways that then improve accessibility to the area. Better access fosters continued urban development, particularly highway-oriented commercial and office land uses. Additional sources of employment come into the area as secondary (and tertiary) industry or services locate near the original major project, spurring on another round of residential development, and so forth.

2. Selection of General Approach

The selection of an approach for testing this theory was tempered by programmatic considerations. The limitations of time and resources precluded highly detailed deterministic approaches patterned after Lowry [5] or others.** They were also inappropriate because of their concern with highly aggregated regional development and/or long planning horizons. Because of this and the problems associated with deterministically modeling a social system, it was decided to use a stochastic approach.

---

* The theory is not entirely new. Most of the urban development models referenced on the same page are based upon the same general relationships posited in our theory. They are usually less explicit than ours, however.

** Forrester [6], Hill [7], Seidman [8], and Center for Real Estate and Urban Economics [9]. In addition to these general or comprehensive models there are many single sector models that have been developed since the early 1960's. See reference [10] and [11] for a review of urban development models.

Resource constraints also made infeasible a dynamic modeling approach because of the effort involved in obtaining longitudinal data to incorporate time into the system. Similarly, a difference approach, modeling the change the development patterns over a time period was infeasible. Consequently, a static approach using the path analytic technique based on regression analysis seemed appropriate to test the theory of induced development.

The static approach to testing the inherently change-oriented theory is justified by three factors: (1) our theoretical assumption that induced development follows a single basic causal structure for all cases or observations, (2) the use of a cross-sectional method of obtaining data for variables observed in a static state and the assumption that input variables were initialized at time zero and held constant long enough so that all the causal consequences in the system were realized, and (3) the use of certain time-lagged exogenous variables in the system. The conceptual usefulness of these factors in testing causal theory is well described in Heise [12] and Blalock [13].

There was not enough existing information on project-level induced development to be able to define the form of the relationships in the system. The form, therefore, was assumed to be linear.* This is not a bad a priori approximation in most social science applications, and it allows the use of well developed statistical techniques. There is accumulating evidence that many social systems can be approximated by a linear function as long as operating conditions remain fairly stable [12]. Even complex nonlinear relationships can often be approximated by a constant relation in discrete subregions of the relationship [14]. Also, if a relationship is thought to be nonlinear on theoretical grounds (e.g., multiplicative, exponential), the data can be transformed prior to entering it in the linear analysis [15].

---

* As part of the testing of the model, a multiplicative relationship was tested with a log transformation. See Chapter VI.

The selection of an appropriate time period was resolved in a somewhat arbitrary manner. Because of the disparate EPA programs GEMLUP focused on and because there was little guidance in the literature, the period of analysis was defined to be ten years.

In summary, the approach consisted of a cross-sectional model that predicts the total land use in the vicinity of a major project ten years after development of a major land use project.

3. Statement of Fundamental Model

The basic theory of induced development may be restated as follows: the amount induced land use is some function of the size of the Major Development Project and certain other variables, viz.,

induced land use = f (major project, other variables)

As indicated previously, the approach used in this study limits one to the use of endogenous variables that measure the total land use at the end of the ten year time period. Conceptually, one can disaggregate the total land use in the area of influence into three components, land use existing prior to the development of the major project, new land use induced by the major project and certain other variables, and new land use not induced by the major project (that is, attributable to some other phenomena such as general regional growth). This may be expressed as:

total land use = prior land use +
project induced land use change +
non project induced land use change

In predicting the total land use in an area of influence, one can identify two types of exogenous variables:

Type I - those used for predicting the induced land use component, such as,

- the size of the major project

- the induced component of the endogenous variables of other land uses
- other independent variables influencing the effect of the major project (i.e., housing vacancy t+0)

Type II - those used for predicting the prior or non-induced land use component,

- the prior and non-induced component of the endogenous land use variables
- those used for predicting the prior land use component (i.e., 1960 housing density)
- those used for predicting the non-induced component (i.e., regional population growth)

Accordingly, the fundamental model that we have assumed is:

$$\begin{aligned} \text{land use}_{t+10} &\equiv \text{prior}_{t+0} + \text{induced}_{t+0 \rightarrow t+10} \\ &\quad + \text{non-induced}_{t+0 \rightarrow t+10} \\ &= f\ (\text{Type I variables, Type II variables}). \end{aligned}$$

We note that both the distinctions between the three land use components and the two variable types are unavailable to this study. The three land use components are not measurable; also, several of our independent variables are possibly of both types.

II. SUMMARY OF STATISTICAL TECHNIQUES (PATH ANALYSIS)

In this chapter we review briefly the concepts, assumptions, and problems associated with the application of path analysis. With regard to the characteristic problems involved in path analysis, we shall focus on those aspects which are most relevant to the project under consideration.

A. AIM OF PATH ANALYSIS

Path analysis was developed by Sewall Wright [16,17,18,19] as a technique for studying and dealing with observed interrelated variables for which it can be assumed that there are several "ultimate" or exogenous variables that completely determine or cause them [20]. It is a method for studying the direct and indirect effects of variables taken as causes of other variables considered as effects [21]. It is not capable of deducing causal relations from available quantitative information (viz., correlation coefficients), but rather it is intended to combine this quantitative information with qualitative interpretation [16]. It is a technique useful in testing theories rather than in generating them, and it can be used to study the logical consequences of various hypotheses involving causal relations. In order to implement the technique, the researcher must make explicit a theoretical framework or model.

B. PATH DIAGRAMS

Figure 2-1 presents various path diagrams which display graphically typical patterns of causal relations among sets of variables. Figure 2-1a shows the relation between seven variables, X_1, X_2, X_3, X_4, U, V, and W. Where X_1, X_2, X_3, X_4 are variables of the systems which are measurable. The variables U, V, and W are unmeasured and outside the system, but do affect variables in the system. These latter variables are often referred to as disturbances or residuals. Usually they are not represented in a path diagram and Figure 2-1a and Figure 2-1b can be considered the same path diagram. The residuals are "understood" to exist. The arrows in all the figures indicate the causal flow. Variable X_1 in Figure 2-1a and 2-1b

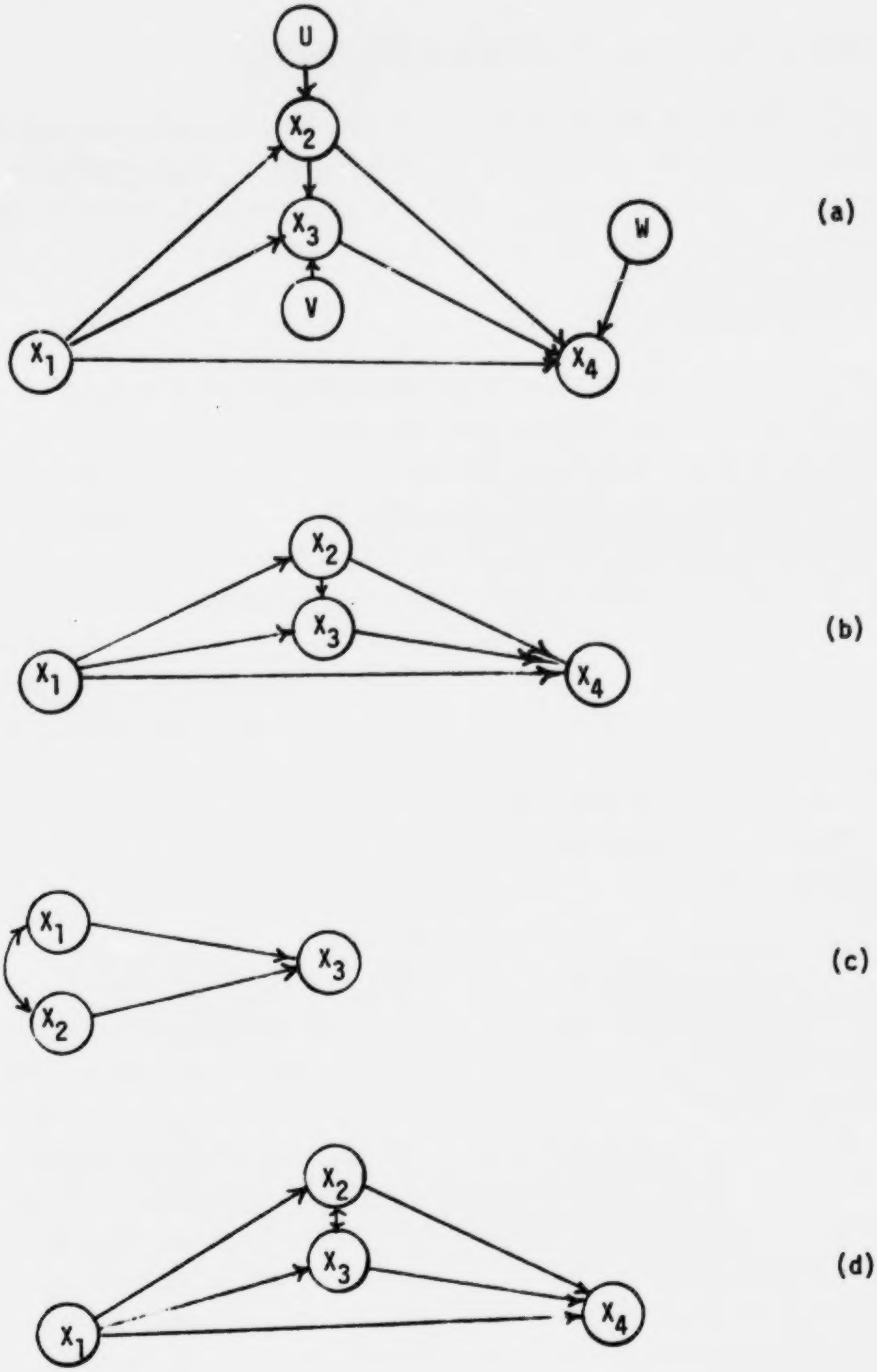

U
X2
X3
V
W
X1
X4
(a)
X2
X3
X1
X4
(b)
X1
X2
X3
(c)
X2
X3
X1
X4
(d)

FIGURE 2-1 ILLUSTRATIVE PATH DIAGRAMS

is called an exogenous variable for its variability and is assumed to be determined by causes outside the system. Variables X_2, X_3, and X_4 of Figures 2-1a and 2-1b are called endogenous. Their variability is explained by exogenous or endogenous variables in the system. Variables U, V and W of Figure 2-1a can be considered exogenous.

Sometimes there is more than one exogenous variable in the system (see Figure 2-1c and variables X_1 and X_2). In this case, the correlation between the two is represented by a curve line. The relation between the exogenous variables is usually not analyzed in path analysis.

The path diagram shown in Figure 2-1d represents an interaction between two endogenous variables X_2 and X_3. Analysis of this type of system presents some serious technical problems which we will discuss later on.* This interaction is also termed a reciprocal causation or a feedback loop.

C. APPROACH (STRUCTURAL MODEL)

In order to analyze the system of variables under investigation the path diagram must be formalized mathematically. A structural model must be developed which defines a causal structure specifying the network of causal paths that exist between variables and identifying the parameters of causation. For Figure 2-1a, the mathematical model would be:

$$\begin{aligned} X_2 &= b_{21}X_1 + U \\ X_3 &= b_{31}X_1 + b_{32}X_2 + V \\ X_4 &= b_{41}X_1 + b_{42}X_2 + b_{43}X_3 + W \end{aligned} \qquad (2\text{-}1)$$

These equations are called structural equations. If the variables are standardized (i.e., have mean zero and standard deviation unity), then the b's are called path coefficients. If they are not standardized, the b's are called path regression coefficients or structural coefficients (there would then also be an intercept in each equation of (2-1)). The b's are the parameters quantifying the causal effects.

---

* See discussion of identification in Section III.D.

Given a structural model, one has an explicit and quanitative statement of theory and so it can be used to explain why variables vary together as they do. It is possible with such a model to calculate how a change in any one variable in the system will affect the values of the other variables. Also, it is possible to analyze how changing the structure of the system would affect the character of the system [22].

D. CONNECTION WITH REGRESSION

It should be emphasized that the three equations of system (2-1) each separately represents a multiple regression equation. The estimates of the b's are obtained using standard least squares (i.e., the standard regression technique). Up to this point, there is no difference between path analysis and standard multiple regression. The b's are beta weights if the X's are standardized, otherwise they are partial regression coefficients.

E. ASSUMPTIONS AND DEVIATIONS

The transition from the path diagram of Figure 2-1a to the system of structural equations (2-1) and the appropriateness of standard least squares (i.e., multiple regression) for estimating the b's of (2-1) involve a number of very important assumptions. We now consider these.

1. Linearity

The change in one variable is assumed to always occur as a linear function of changes in other variables. If it is believed or known that the relation is nonlinear, for example, say multiplicative:

$$X_3 = e^{(b_{31}X_1 + b_{32}X_2 + U)} \tag{2-2}$$

then the data must be transformed before the analysis. If there is nonlinear "causality" the usual path analysis will not detect it.

2. Recursive

In order for the above discussion to be correct, the system must not contain reciprocal causation or feedback loops (e.g., as in Figure 2-1d with the variables X_2 and X_3). The causal flow must be unidirectional. At a given point in time, a variable cannot be both a cause and an effect of another variable. Such a system is called recursive.

It is possible to solve a system of structural equations if there are feedback loops by use of instrumental variables. Instrumental variables are exogenous variables introduced into the system for this specific purpose. They are used routinely in economic analysis [23,24]. Basically what they do is to put enough constraints on the structural system so that by use of estimation techniques (either two-stage least squares or limited information least-generalized residual-variance solution) [23] the path coefficients can be estimated. A point often not emphasized in the "traditional" sources of path analysis is that little is known about the resulting estimates of the path coefficients. The most one can say is that they are consistent estimates of the true b's. This means that they converge in probability to the true b's or, in other words, as the sample size increases they get "closer" to the true b's. Basically, nothing is known about the estimates for small samples (such as 20 in our present problem). All the "nice" properties of the multiple regression estimates (i.e. minimum variance, unbiased estimates) do not hold for the new ones.

In the present project we had to make use of the above techniques (i.e., use of instrumental variables and two stage least squares) to solve for the b's. In order to judge the usefulness and appropriateness of the estimated b's we used four approaches simultaneously. The first three approaches involved the use of the approximate t tests (to judge if the b's are statistically significantly different from zero), the beta weights (to judge if in the standardized scale the magnitude of the effect

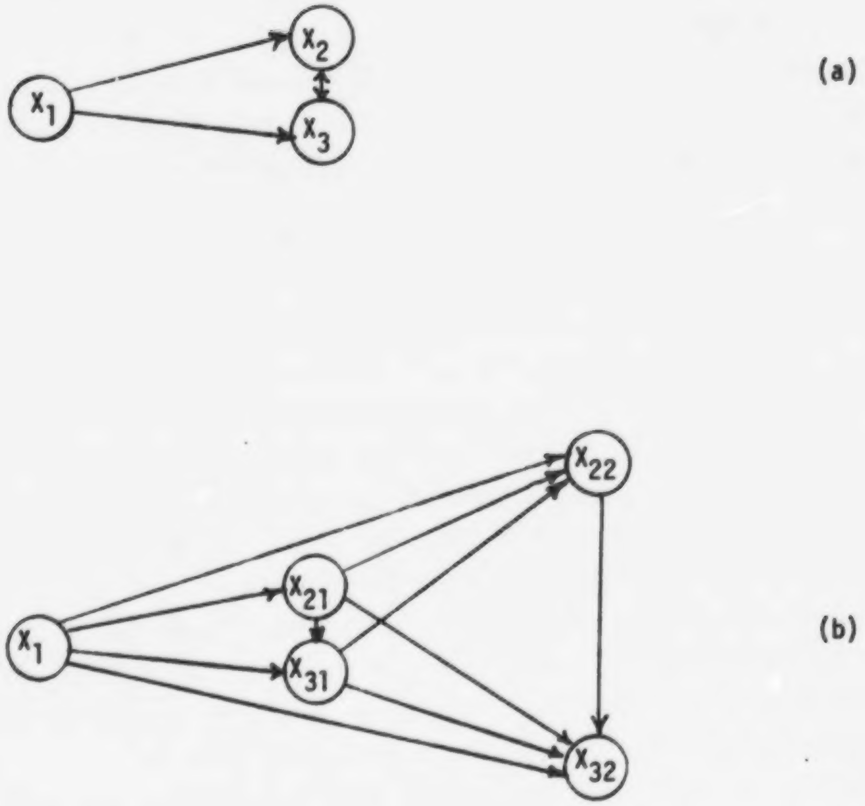

X_2
X_1
X_3
(a)
X_{22}
X_{21}
X_1
X_{31}
(b)
X_{32}

FIGURE 2-2 LEAD - LAG RELATIONSHIPS

is "large") and the signs of the b's (to verify that the signs of the b's are what theory would predict). The fourth approach involved the use of "Jackknifing" to verify that the b's are stable (i.e., will not change radically with the inclusion or exclusion of just one observation.*) The actual implementation of the four approaches is discussed in detail in Chapter VI on analysis.

Another way of solving the system if feedback loops exist is to use lead-lag relations. This might be possible in longitudinal data [19]. Consider the path diagrams of Figure 2-2. In Figure 2-2a, we have an interaction between variables X_2 and X_3. However, say X_2 and X_3 represent the final states after a time period (say, 10 years, for example). Further, say it is possible to break the 10 years into two time periods (time periods 1 and 2) (see Figure 2-2b). In time period 1, we may have variable X_2 (called X_{21}) as a cause of X_3 (called X_{31}) but not X_{31} as a cause of X_{21}. Then into time period 2 we have X_3 as a cause of X_2 (i.e., X_{31} X_{22}). The system shown in Figure 2-2b is recursive and given sufficient data it is solvable. The dynamics of the system over the full time period are needed. Unfortunately, for the present study sufficient data was not available to implement this lead-lag device.

3. Causal Priorities

The causal priorities among the variables must be "correctly" established outside of the analysis. A major drawback of path analysis is that it cannot establish causal relationships using data alone. One needs as inputs both data and theory. The ordering of the variables (that is, the direction of the arrows in the path diagrams) must be established by the investigator. "There is no error-check mechanism in path analysis to reject an incorrect ordering" [22]. It is common that the data can fit well two very different orderings. The selection of the appropriate one depends on correct theory. When one is dealing with variables ordered in time it is often possible to know the causal priorities. Path analysis seems well suited for this type of problem. For the present study, the ordering in time was usually not available so a substantial effort involving substantive considerations was needed to aid in establishing the causal priorities. Also, by use of the instrumental variables and two stage least squares for estimating the b's in feedback loops, the causal priorities of some questionable situations were established (e.g., for the cases where it is not clear that a causal priority exists if the estimated b for a direction (say X_1 to X_2) is zero then the analysis indicates there is no causal priority of X_1 to X_2).

---

* The concept of "Jackknifing" is discussed further in Part 5 of this Chapter.

4. Uncorrelated Residuals of Endogenous Variables

In the actual analysis the observed correlations among the observed variables are used to estimate or identify the path coefficients. If the residual (i.e., the disturbances) are uncorrelated, then the usual least squares technique (i.e., multiple regression) can be used for the estimation. If, however, the residuals are correlated, then the path coefficient cannot be identified by the ordinary least squares technique for there are then too many unknowns. An implication of this assumption of uncorrelated residuals is that all inputs must be entered explicitly into the analysis. (It is possible to ignore a variable that affects two or more inputs as long as it does not affect the residuals of any endogenous variables.) This restriction of uncorrelated residuals is very important and meaningless and confusing results may come out of the analysis if it is not met.

Because of its importance to the project under consideration, we shall develop a bit further here the consequences of not including all inputs explicitly in the model. Say we have three variables X_1, X_2, and X_3 which are causal factors of a dependent variable Y. The appropriate equation is:

$$Y = b_{y1}X_1 + b_{y2}X_2 + b_{y3}X_3 + e \qquad (2\text{-}3)$$

Here, e is the residual. Now say we do not include X_3 in our system and we "believe" the correct equation is:

$$Y = b_{y1}X_1 + b_{y2}X_2 + e' \qquad (2-4)$$

e' in (2-4) is a new residual. It is actually $e+b_{y3}X_3$. It can be shown [23] that the estimate we will obtain for b_{y1} will actually be an estimate not of b_{y1} alone, but rather of

$$b_{y1} + b_{y3}\,(B_{31.2}) \qquad (2-5)$$

Where b_{y3} is the real path coefficient given in (2-3) and $B_{31.2}$ is the regression coefficient of X_3 and X_1 after allowing for the effects of X_2. The important point here is that if we leave out an input it will reappear in our path coefficients. We will have no way of knowing how much of our estimate is the real b_{y1} and how much is $b_{y3}(B_{31.2})$. It is possible that b_{y1} is positive yet the value of (2-5) is negative (a complete change in sign). This point is not made very clear in much of the literature in path analysis.

Another implication of this fourth assumption is that all residuals should be uncorrelated with all other exogenous variables that precede them in the system. If one suspects that the residuals are correlated, then the technique of two stage least squares can be used to remove this problem. However, in doing so, it introduces the problem of having estimates whose statistical properties are not well known (see part 2 above). In the present project, for those equations not requiring two stage least squares estimation techniques, we assumed that the models contain a sufficient number of relevant variables to minimize this problem of correlated residuals.

5. Stability of the Estimates

In order for path analysis to be of use the resulting estimates should be stable. They will be of little use if they change radically, depending on the sample observations. In most applications of path analysis this is achieved by using large samples and reliable measuring instruments. Tukey [26] and Turner and Stevens [27] with regard to this point suggested using path regression coefficients rather than the usual path coefficients (i.e., beta weights). Path regression coefficients tend to be more stable as one applies the same structural model to different populations.

For the problem of this project we have only 20 observations to do all the required estimation. This is a very small sample. There will be a real problem of obtaining stable estimates. One way of judging the stability of the estimates is to use the statistical technique of "Jackknife". Instead of running one regression using 20 observations, 20 regressions each containing nineteen observations are run. The path coefficients can then be examined for stability. This technique will not allow the path analysis to capitalize on chance elements in the data. This Jackknifing was done for a selected number of equations in the present project.

6. Multicollinearity not Present

If the correlations between independent variables in a multiple regression are extremely large in absolute value, we have what is called the problem of multicollinearity. If present it is usually impossible to separate the influences of the variables involved. The effect on the path analysis will be to produce unstable path coefficients (they will have large standard errors). Beta weights greater than unity are even possible. Obviously, the purposes of path analysis can be defeated if collinear variables are present. Notice these collinear variables can be both exogenous and endogenous as long as they appear somewhere as independent variables.

To handle the problem of multicollinearity we examined the zero order correlations for large correlations and paid careful attention to the output of individual regressions looking for unusually large (in absolute value) beta weights and standard errors of the regression coefficients. See Chapter VI on analysis for further details.

7. Scale

At one time it was considered important that the variables be measured on an interval scale. Now, however, this is not so important if properly handled [28]. The data for the present proposal does appear to be appropriate for path analysis.

8. Suppressor Variable Problem

It is possible for a path analysis system to contain too many input variables. For example, say X_1 refers to the size of a new highway and X_4 is the commercial land use. Further, say X_2 and X_3 are two other variables. The structural model is given by the equations:

$$\begin{aligned} X_2 &= b_{21}X_1 + b_{2u}U \\ X_3 &= b_{31}X_1 + b_{3v}V \qquad (2\text{-}6) \\ X_4 &= b_{41}X_1 + b_{42}X_2 + b_{43}X_3 + b_{4w}W \end{aligned}$$

and say the correlation matrix is:

	x_1	x_2	x_3	x_4
x_1	1.000	.800	.600	.000
x_2		1.000	.430	.288
x_3			1.000	.384
x_4				1.000

Given this it can be shown that the solution for X_4 is:

$$X_4 = -X_1 + .8X_2 + .6X_3 + .736W \qquad (2\text{-}7)$$

The size of the highway appears to have a negative impact on commercial land use (path coefficient is -1).

The actual situation is that a model was developed to produce the correlation matrix and it was:

$$X_4 = .48U + .48V + .736W \qquad (2\text{-}8)$$

U, V and W are the residuals of X_2, X_3 and X_4. X_4 and X_1 are actually uncorrelated. The "strange" result shown in (2-7) arises because of the following: If we use X_2 and X_3 to predict X_4 we will have a useful result because X_2 and X_3 contain X_1 which is irrelevant. In (2-7) the term X_1 is used to "suppress" the irrelevant components in X_2 and X_3. That is, X_2 and X_3 do contain a portion of X_1 in them and this portion is subtracted again in the equation by adding the negative term with X_1. Figure 2-3 illustrates the problem. The system of equation (2-6) is assuming Figure 2-3a while the true situation is Figure 2-3b.

The importance of this illustration is that one cannot just include anything into a structural system and believe all will work out well.

To handle this problem we examined the zero order correlations between all endogenous variables (i.e., those that are dependent variables in some regressions) with all other variables. Those cases where zero or near zero correlations appeared were noted. Later if the partial relations (when included in multiple regression or two stage least squares) proved to be significant, careful substantive consideration was applied to see if a suppressor variable problem may exist. Also the intercorrelations of the regression coefficients or beta weights (i.e., the b's) were examined to detect if the suppressor variable problem existed. See Chapter VI on analysis for further details.

F. OTHER ASSUMPTIONS

There are other assumptions that deal solely with the fact that path analysis is a regression technique (e.g., independence of observations). We will not enumerate them for they are standard [25].

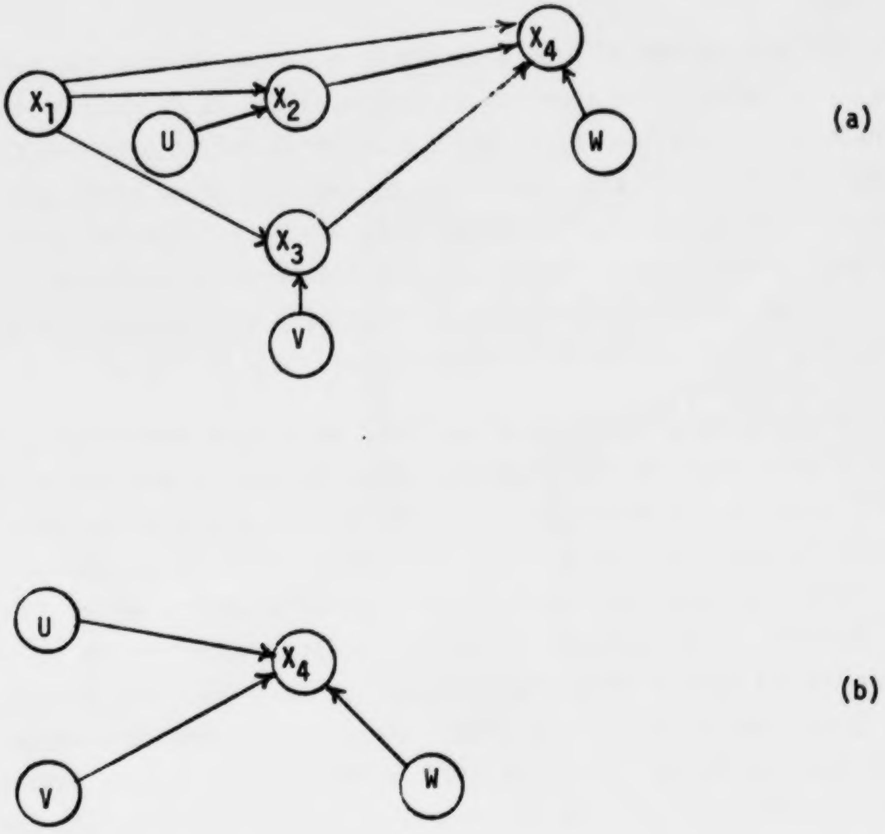

X_1
X_2
X_3
X_4
U
V
W
(a)
U
X_4
V
W
(b)

FIGURE 2-3 ILLUSTRATION OF SUPPRESSOR VARIABLE

G. THEORY TRIMMING

To end this review of path analysis we will now discuss the question of theory trimming. One approach to path analysis is to develop the most elaborate system and then after estimating the involved path coefficients attempt to refine or trim the system by dropping those paths that have coefficients "close to zero". Those paths that are "close to zero" imply no or only a weak causal linkage between the involved variables. The possibility for refining or trimming a theory and thus making a theory more parsimonious is of considerable significance.

There are a number of ways of deciding which path coefficients can be dropped. First, there is the criterion of statistical significance (e.g., drop all path ceofficients which do not differ significantly in the statistical sense (0.05 level of significance) from zero. With large samples, this could imply retaining very small coefficients and with small samples this would imply dropping very large coefficients. Second, there is the possibility of dropping all coefficients whose values are less than some preassigned fixed value (e.g., say 0.05 or 0.03) [29]. There is also the dependence analysis procedure of Boudon [30]. In this technique, one usually assumes that some path coefficients of the structural equations are zero. To test this procedure, one sees how well the non-zero paths reproduce the original correlation matrix. In general, for any trimming procedure it is important to see how deletion of some path coefficients affects the ability to reproduce the original observed correlations (an important property of path analysis itself is the ability of reproducing the original observed correlations by means of the path coefficients).

The approach we took involved retaining a variable if (1) its t value exceeded unity in absolute value (this guarantees that the adjusted R is larger with its inclusion than with its exclusion for ordinary multiple regression), (2) its beta weight exceeded .1 in absolute value (this was judged to reflect a substantively meaningful relation) and (3) if it was deemed a priori to be of substantive importance and its sign (i.e., the sign of the coefficient b) was in the right direction. See Chapter VI on analysis for further detail.

III. INITIAL LAND USE MODEL SPECIFICATION

A. OBJECTIVE

The objective of this phase of the project was to specify an initial land use model explaining induced or associated land use ten years after the construction and operation of a major project. The basic theory underlying the development of the model was that major projects have certain associated or induced land uses and these land uses can be predicted based on the characteristics of the major project and the area in which it locates. However, because of the approach to testing this model, it is necessary to include all land uses in the vicinity of the major project, whether they were induced or not induced, or even existing prior to the construction of the major project.

Two types of major projects were to be considered in the formulation of the model, residential projects and office/industrial projects. Because of the differing land uses associated or induced by these types of projects, a separate model was constructed for each. Thus, two models were developed, one explaining induced or associated land uses ten years after the construction and operation of a major residential project; the other explaining induced or associated land uses ten years after the construction and operation of a major industrial/office project.

B. DEFINITION OF MAJOR PROJECT

For purposes of the model specification, a major residential project was defined as housing facilities, planned unit developments or new towns containing a minimum population of 4,500; a major industrial/office project was defined as an office or industrial park or a research and development complex with a minimum employment of 2,250. Both types of projects were initially assumed to reach nearly 80 percent occupancy within two years of operation. However, during case study selection, the definition of major project was somewhat modified to permit phased projects (See page 4-3).

In addition, for purposes of calibrating the model, the case studies to be analyzed were required to be projects built between 1954 and 1964. The induced or associated land uses were those as of the year 1970; i.e. the year by which it was assumed that the land use impacts of the project had stabilized.

C. AREA OF INFLUENCE

The size and definition of the geographical area for which the causal model and predictive equations are specified and calibrated is critical to their success and applicability. If the area selected is too large, the model will accurately predict the end state land use, but the influence of the major project will not be seen. This is due to the noninduced land use overwhelming the induced land use. On the other hand, if the area selected is too small, the model will not capture all of the induced uses; this also leads to an underestimate of the major project's influence. Consequently, two issues concerning the area of influence required resolution:

1. Should the area of influence vary with the size or density of the major project?

The statistical analysis to be employed requires that the major projects show substantial variation in size. This assumes that a larger project will have a greater amount of induced and secondary effects.

As a major project increases in size or decreases in density it will occupy a relatively larger amount of land. The more land area in the major project, the less area there will be in the remaining circle. Figure 3-1(a) illustrates this concept. With a larger amount of land area in the major project, a circle of radius r' would have to be drawn so that the area inside the circle, but not in the major project, remains constant. Few of the major projects considered as case studies were much more than one square mile. This is a relatively small amount of land in a circle of radius two miles (3.22 km) and area of 12.56 square miles (32.53 sq. km) or a circle of radius 3 miles (4.83 km) and area of 28.27 square miles (73.22 sq.km). Therefore, it is felt that this concern can be ignored.

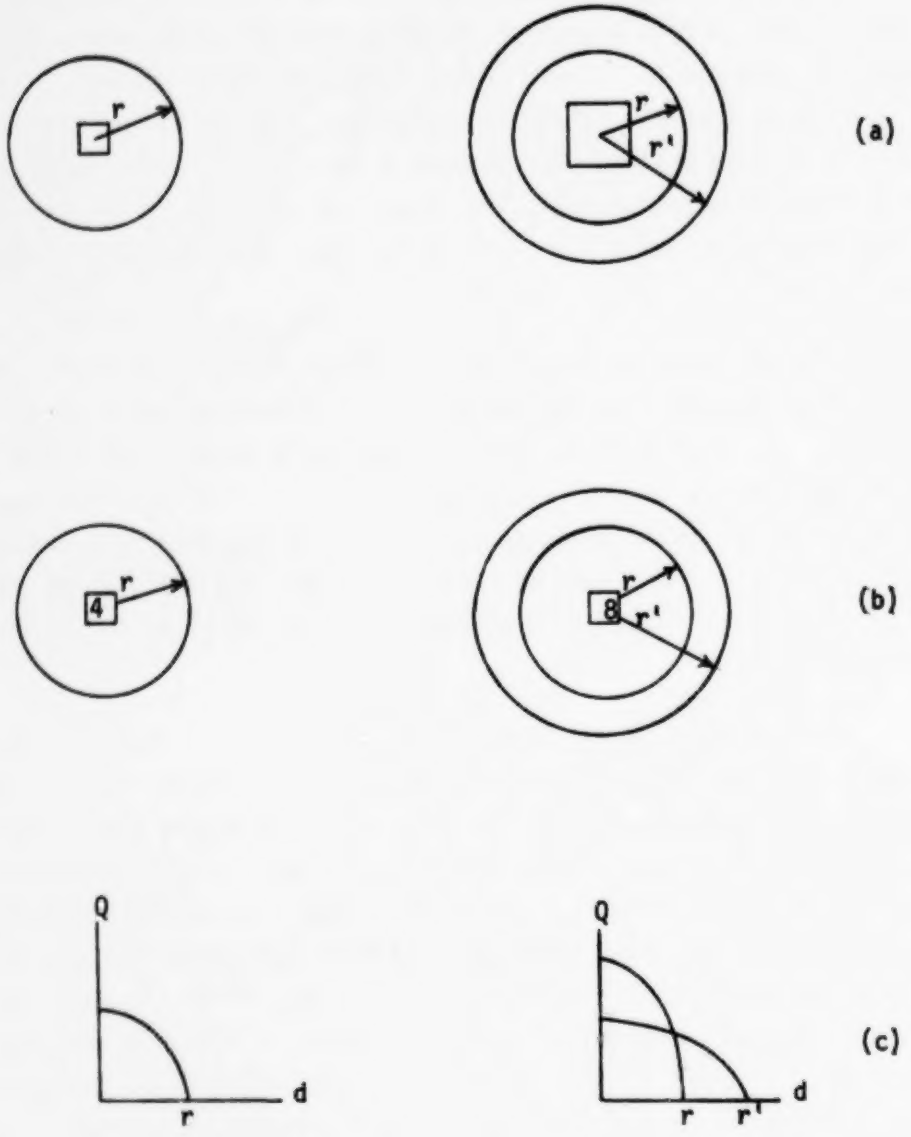

r
r
r'
(a)
4
r
8
r
r'
(b)
Q
d
r
Q
d
r
r'
(c)

FIGURE 3-1 ILLUSTRATION OF EFFECT OF DIFFERING MAJOR PROJECT SIZES

Even with major projects of identical area, a higher density project is assumed to have a greater inducing effect (due to more employees or residents). Figures 3-1(b) and 3-1(c) illustrate this concept. For a project of size 4, a circle of radius r is drawn, and the amount of induced land use Q is plotted against the distance d from the major project. At distance r from the major project, the induced land use falls to zero. The size of the area under the curve is the total induced land use for a project of size 4.

To the right in Figure 3-1(b) and 3-1(c), a project of size 8 is shown. It is assumed that the project has a larger induced land use, so the area under the curve must be greater. Depending on where this additional induced land use occurs, a circle of radius r or r' would be appropriate. If the curve represented by distance r' is the true case and a circle of radius r is used, some of the induced land use will not be captured, and the statistical analysis will underestimate the impact of the larger major project.

In reality, the curve for a larger project is somewhere between the extremes shown in Figure 3-1(c). Although it is believed that a larger major project will have some influence at a greater distance than a smaller project, we have assumed that the additional induced land use between r and r' is negligible. This assumption was made in the interest of simplifying the model. If a varying study area (area of influence) were allowed, it would create two difficult problems beyond the scope of this study: 1) the development of a systematic definition of the area of influence as a function of major project size and 2) the definition of the induced land uses as a rate, i.e., square feet of induced land use per acre of study area.

In conclusion, the size of the area of influence will be constant, regardless of major project size.

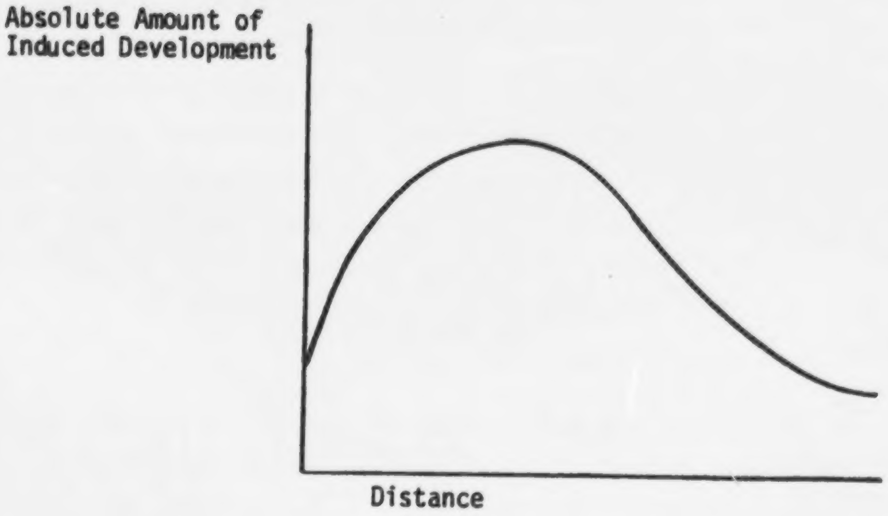
2. What is the appropriate size of the study area (area of influence) for each major project type?

The induced growth from a major development project would be distributed throughout an urban area; however, in selecting the size of the area of influence, we have assumed that the absolute amount of induced development is some declining function of the distance from the major project. The absolute amount of induced development will, of course, be influenced by other variables (the determination of that influence is the goal of this project.

For example, the construction of a major residential development would create a demand for retail trade facilities to serve the residents. The residents would generally minimize travel time and would shop at shopping centers closest to them. Sales due to the residents of the major development would, therefore, decline with the distance from the development. The sales could occur at new shopping centers or at existing ones enlarged to handle the increased volume.

Induced residential development from a major industrial or office development would generally follow the same pattern. The employees would seek to minimize their travel time from home to work while trading off other factors such as the cost of land, quality of the residential neighborhood, etc.

At a small distance to the major project, the absolute amount of land available for development would be limited. Accordingly, the first assumption is modified such that the absolute amount of induced development is at first an increasing function with distance at small distances from the major development, viz.,



The general shape of this curve is confirmed by the sole item of literature relevant to this question known to us. The Air Pollution Impact Methodology for Airports - Phase I, [31], in part, analyzed changes in land use before and after the development of Chicago's O'Hare Airport in the early 60's. Figure 3-2, reproduced from that study, shows the relative proportion of land devoted to a particular use by distance from the airport and time from development. The horizontal axis, labeled zone, refers to 5 concentric squares surrounding the airport (see Figure 3-3), each one is approximately an additional mile (1.609 kilometers) from the airport. The vertical axis, labeled "percent zone land use", refers to the percent of the land in each zone that is devoted to a particular use.

Manufacturing and warehousing land use would seem to be the most significant induced use from the airport development. The percentage of land that is industrial in 1970 and 1960 is tabulated on the following page.

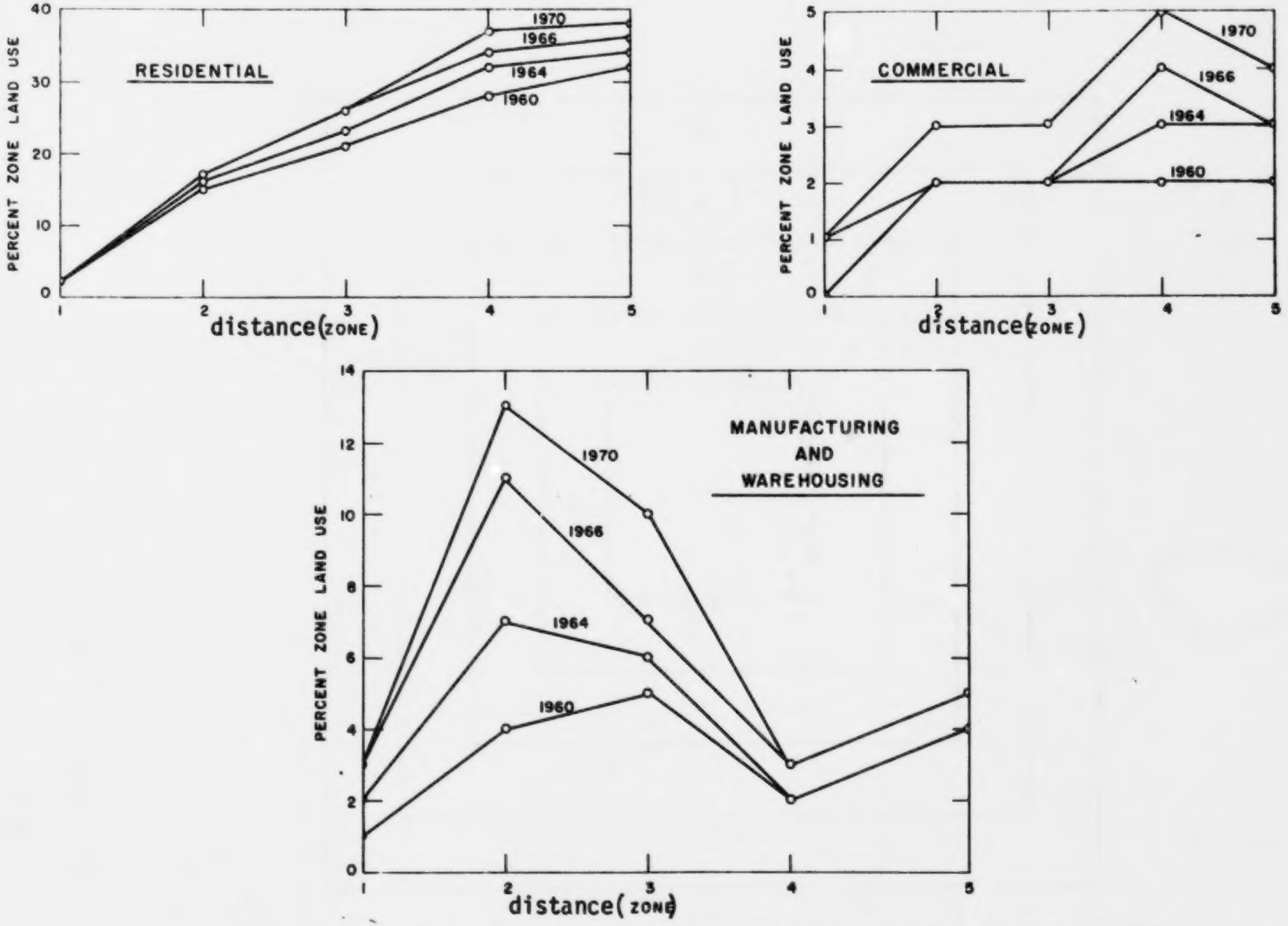


FIGURE 3-2 RELATIONSHIPS OF LAND USE TO DISTANCE FROM THE AIRPORT

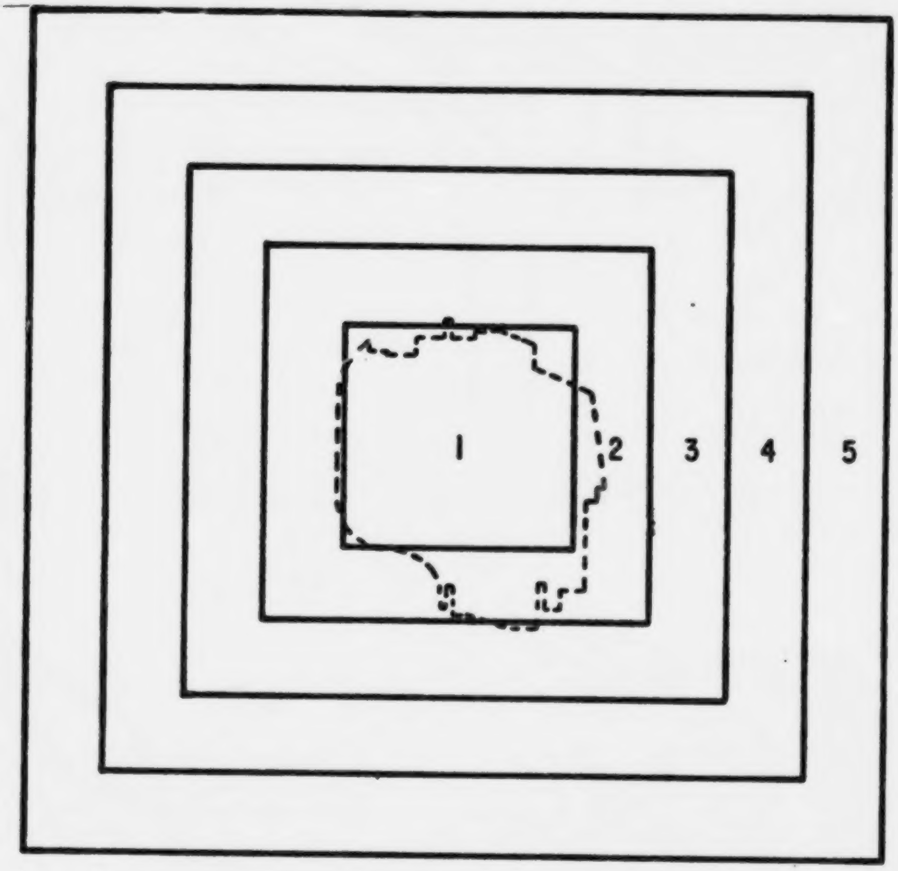


FIGURE 3-3 ZONES USED FOR LAND USE/DISTANCE ANALYSIS

NOTE: The Dotted Line Indicates the Boundaries of the Airport.

Zone	1970 Industrial Percentage	1960 Industrial Percentage	60-70 Percentage Growth
1	3%	1%	200
2	13%	4%	223
3	10%	5%	100
4	3%	2%	50
5	5%	4%	25

Both the graph (Figure 3-2) and the data above seem to support the general shape of the curve if one assumes the change in industrial land use around O'Hare Airport is solely attributable to the airport.

One criterion for selecting the size of the study area would then be to maximize the amount of the induced development that would be included. In the case of an airport, we conclude that three to four miles would be sufficient, however, the land use model being developed under the current contract will predict the total land use rather than the change in land use patterns induced by a major development. The model is cross-sectional; thus, the successful identification of the impacts of a major project will depend on observing a correlation between the frequency of a particular land use and the size of a major project. For example, if the real state of the world is that airports induce industrial development, then a study design using cross-sectional analysis and a small sample size will have the best changes of identifying the real state of the world if the study area is selected such that the ratio of induced (1960-1970) industrial land uses to existing (1960) land uses is maximized.

This ratio is unknown since identifying induced land uses is the goal of the contract effort. It is also probable that the ratio will vary with the type of land use being predicted, the size of the major project, and the category of the major project. With these caveats, however, an indication of the appropriate study area size can be obtained.

a. Induced Residential Example

The national average for home to work trip lengths for industrial plants [32] provide a rough estimate of the location of the induced residential development. Using Middlesex County, Massachusetts, as an example, the ratio of induced residential development to existing housing is computed in Exhibit 3-1.

At any radius that is greater than two miles, the number of induced households as a percentage of existing households falls below 10 percent. Given the small sample size in this study, it is doubtful the analysis could be successful at a greater radius. On the other hand, a radius of two miles or less would compromise the usefulness of the product as less than half (800 households/2000 households =40%) of the total induced households would be in the study area.

b. Induced Commercial Example

Using a methodology similar to the above, the location of induced commercial development can be obtained. As shown on Exhibit 3-2, the multiplier dy/dx, measuring the increase in regional income, Y, per increase in export, X, is roughly estimated to be 1.45. This approach treats the development of a major residential project, with its 2000 household consumption of goods and services, as an export as far as the region is concerned. The increase in regional goods and services due to 2000 household major projects is then estimated. Finally, the ratio of induced to prior commercial development is estimated. As indicated, it peaks at 2 miles; however, note that a circle of this radius includes only 30 percent of the total induced commercial development.

In summary, the three examples indicate that a radius of between 2 and 3 miles would be appropriate. We have elected to use a study area of 10,000 acres, i.e., a circle with a radius of 2.23 miles.

EXHIBIT 3-1

INDUCED RESIDENTIAL EXAMPLE

Objective: Estimate of the induced residential development as a percentage of prior residential.

Assume a plane of constant household density with the subsequent introduction of a 2000 employee industrial facility. Middlesex County, Massachusetts, will be used as an example.

I. DATA

Middlesex County:

Households = 461,000 [33]
Total Acres = 540,564 [33]
Household Density = 0.8528 households per acre

National home to work trip length frequency of auto drivers and walks to work [32]:

Distances		Cumulative Percentage of Auto Drivers and Walk to Work Commuters
Kilometers	Miles	
.805	.5	5.0
1.609	1.0	17.5
3.218	2.0	40.0
4.827	3.0	61.5
6.436	4.0	68.5

EXHIBIT 3-1 (Continued)

INDUCED RESIDENTIAL EXAMPLE

Radius Miles 1	Km 2	Area (πR^2 * 640) 3		Cumulative Prior Households (#3 * 0.85) 4	Cumulative Proportion of Induced Households 5	Cumulative as Percent Households (#4 * 2000)	Induced Induced of Prior (#5/#3)
		Acres	Sq. Meter				
.5	.8	503	$2.0\text{x}10^6$	429	0.05	100	23.3
1.0	1.6	2011	$8.1\text{x}10^6$	1709	0.175	350	20.5
2.0	3.2	8042	$3.3\text{x}10^7$	6836	0.40	800	11.7
3.0	4.8	18096	$7.3\text{x}10^7$	15381	0.615	1230	8.0
4.0	6.4	32169	$1.3\text{x}10^8$	27344	0.685	1370	5.0

EXHIBIT 3-2
INDUCED COMMERCIAL DEVELOPMENT

Objective: Estimate the induced retail trade and services as a percentage of total retail trade and services.

Middlesex County, Massachusetts, is used as an example.

I. Data

Middlesex County:

Retail Trade, Thousand Dollars = 3,703,001 [33]
Services, Thousand Dollars = 1,090,370 [33]
TOTAL = 4,793,371
8,867 Thousand Dollars of Commercial Sales per Acre
10,398 Thousand Dollars of Commercial Sales per Household
Median Disposable Income = 13076 Dollars [34]
Disposable Income = 6,782,652 Thousand Dollars

II. Income Equation

Conceptually, the major project can be thought of as an export.

$$Y = C_{l_g} + C_{l_s} - M_g - M_s + I + G + X$$

where

Y = regional income
C_{l_g} = consumption of local goods
C_{l_s} = consumption of local service
M_g = imports of goods
M_s = imports of services
I = investment
G = government
X = exports

EXHIBIT 3-2 (CONTINUED)
INDUCED COMMERCIAL DEVELOPMENT

Consumption Local Goods: $C_{1_g} = 3703001/6782652 = .54595$ .54595

Consumption Local Services: $C_{1_s} = 1090370/6782652 = .16076$ $\underline{.16076}$

TOTAL .70671

Assume 60% of goods imported

Assume 40% of services imported

$$Y = .54595Y + .16076Y - .6(.54595)Y - .4(.16076)Y + I + G + X$$

$$Y = (I + G + X)/.685$$

$$dY/dx = 1.45$$

III. Induced Land Use Calculation

$$\Delta X = \Delta \text{Households} * \$13{,}076$$

$$\Delta Y = \Delta X * 1.45$$

$$\Delta \text{Commercial} = \Delta Y * 0.70671$$

Assume that prior retail trade and services evenly distributed, 8.867 thousand dollars per acre. Assume a 2000 household major project.

Shopping center trip length distribution [32].

	distance	
Miles	KM	
1	1.6	5%
2	3.2	30%
3	4.8	50%
4	6.4	60%
5	8.0	75%

EXHIBIT 3-2 (CONTINUED)

INDUCED COMMERCIAL DEVELOPMENT

Radius (miles)	Radius (Km)	Area (acres)	Area (m^2)	Prior Commercial Activity (Dollars)	Cumulative Trip Frequency	Households Shopping in Area	ΔX	ΔY	Induced Commercial (Dollars)	Induced Commercial / Prior Commercial
1	1.6	2011	$8.1\text{x}10^6$	17831	.05	100	1307.6	1896	1340	7.5%
2	3.2	8042	$3.3\text{x}10^7$	71308	.30	600	7845	11376	8039	11.3%
3	4.8	18096	$7.3\text{x}10^7$	160457	.50	1000	13076	18096	13399	8.3%
4	6.4	32169	$1.3\text{x}10^8$	285243	.60	1200	15691	22752	16079	5.6%
5	8.0	50265	$2.0\text{x}10^8$	445703	.70	1500	19614	28440	20099	4.5%

D. IDENTIFICATION

In this section we briefly review the limitations imposed on the model specification due to the requirement of identification.

When there are feedback loops or reciprocal causal paths in a path analytic model, then the residuals in the model are correlated with the endogenous variables and ordinary least squares yields biased and inconsistent estimates of the structural coefficients. For example, say X_1 and X_2 are residential and commercial variables measured in appropriate units and say there is believed to be a feedback loop between these. Then possible structural equations are:

$$\begin{aligned} Y_1 &= a_1 + b_{21}Y_2X_2 + e_1 \\ Y_2 &= a_2 + b_{12}Y_1X_1 + e_2 \end{aligned} \qquad (3\text{-}1)$$

It can be shown [35] that the residual e_1 is correlated with Y_2 and the residual e_2 is correlated with Y_1. Ordinary least squares applied to either of these equations will result in biased and inconsistent estimates of the structural parameters a_1, b_{21}, a_2, and b_{12}. By biased and inconsistent estimates we mean that the average value of the estimates will not equal the values of the structural parameters no matter how large the sample size may be.

The approach usually followed at this point to obtain appropriate estimates of the structural parameters is to introduce instrumental variables and apply two stage least squares. This approach will yield consistent estimates if the relevant equations are properly identified. The term "instrumental variable" merely means a variable which acts as an instrument in making the system identifiable. Below, we review how instrumental variables are introduced, how two stage least squares is applied and how the identifiability condition arises.

Given the structural system of equation 3-1, one must introduce at least one new variable which is a "cause" of Y_1 but not of Y_2. Let us call this variable X_3. Similarly a X_4 should be introduced for Y_2. X_3 and X_4 are instrumental variables. There is no reason why there need be only one instrumental variable for each endogenous variable. There must be at least one. To allow for more generality let us say we have a second instrumental variable for Y_1 and let us call it X_5. The important point is that we need some instrumental variables which are causally related to one of the endogenous variables, but not to the other in the system. It is possible to have some exogenous variables (which are still often called instrumental variables) which are causes of more than one of the endogenous variables. For our example, let us say X_6 is such a variable.

The revised structural system is now,

$$\begin{aligned} Y_1 &= a_1 + b_{21}Y_2 + b_{31}X_3 + b_{51}X_5 + b_{61}X_6 + e'_1 \\ Y_2 &= a_2 + b_{12}Y_1 + b_{42}X_4 + b_{62}X_6 + e'_2 \end{aligned} \qquad (3\text{-}2)$$

The e'_1 and e'_2 of (3-2) are not the same residuals as e_1 and e_2 of (3-1).

It can still be shown that Y_2 is correlated with e'_1 and Y_1 is correlated with e'_2 (i.e., we have correlated residuals), so ordinary least squares still cannot be applied directly to obtain consistent estimates of the structural parameter b's. However, if we manipulate the equations of (3-2) a new system of equations which have Y_1 and Y_2 only on the left viz.,

$$\begin{aligned} Y_1 &= a'_1 + b'_{31}X_3 + b'_{41}X_4 + b'_{51}X_5 + b'_{61}X_6 + e''_1 \\ Y_2 &= a'_2 + b'_{32}X_3 + b'_{42}X_4 + b'_{52}X_5 + b'_{62}X_6 + e''_2 \end{aligned} \qquad (3\text{-}3)$$

is possible. This new set of equations gives us a way of estimating the b of (3-2). Notice these equations have the instrumental variables on the right hand side. The equation of (3-3) are often called the reduced form equations. Ordinary least squares may be applied to obtain estimates of the a's and b's (and in turn estimates of Y_1 and Y_2 are possible). This is the first stage in the two stage least squares.

Next, the estimates of Y_1 and Y_2 obtained from (3-3) can be substituted into the right hand side of the equations in (3-2). Using Y_1 and Y_2 to represent the estimates, the system (3-2) with this adjustment becomes,

$$\begin{aligned} Y_1 &= a_1 + b_{21}\hat{Y}_2 + b_{31}X_3 + b_{51}X_5 + b_{61}X_6 + e'_1 \\ Y_2 &= a_2 + b_{12}\hat{Y}_1 + b_{42}X_4 + b_{62}X_6 + e'_2 \end{aligned} \qquad (3\text{-}4)$$

$\hat{Y}_2$ is not correlated with e'_1 and Y_1 is not correlated with e'_2 and so ordinary least squares may now be applied to obtain consistent estimates of the structural parameter, the b's. This is the second stage in the two stage least squares.

In order to obtain consistent estimates of the b's it is necessary that $\hat{Y}_2$ is not a linear function of X_3, X_5 and X_6 alone. This is the identification problem (we want to "identify" or estimate the b's) and by bringing X_4 in the system we have this necessary condition satisfied. Recall X_4 is the instrumental variable that is causally related to Y_2 but not Y_1. A similar necessary condition is needed to identify the b's of the second equation of (3-4).

The necessary condition for identification can be formalized as follows: Say we want to estimate the b's of the first equation of the system (3-2). First, one counts the number of instrumental variables in the full system. Here, there are 4 - X_3, X_5, X_6 and X_4. Next, one counts the number of instrumental variables excluded from the equation under consideration. Let us call this number m_0. Here m_0 is one for variable X_4. Then we count the number of endogenous variables on the right hand side of the equation. Let us call this number q. q for this problem is one for Y_2. The necessary condition for identifying the parameters of the equation under consideration is then, that,

$$m_0 \geq q - 1 \qquad (3\text{-}5)$$

[35]. For our example $1 \geq 1 - 1 = 0$ so we have identification. Condition (3-5) is necessary but not sufficient (i.e., if an equation is identified it must satisfy (3-5), however, there are equations that satisfy (3-5) but are not identified.) The sufficient condition involves rank conditions of various matrices [35].

Due to the existence of feedback loops in the present project, the problem of identification required consideration. In the model specification phase of this project, care was taken to specify exogenous variables that could serve as instrumental variables. Additionally, the necessary condition for identification (3-5) was continually checked for compliance.

During the analysis phase of this project (Chapter VI) an unanticipated problem related to identification developed. As indicated above, in the first stage of the two stage least squares, ordinary least squares is applied to the right hand side endogenous variables. All the instrumental variables in the system are used as right hand side variables. As the model, as specified later in this chapter, had five simultaneous equations, there were numerous instrumental variables. In some instances, they approached twenty, which is the sample size. Thus, the degrees of freedom approached zero. This problem and its resolution is discussed in the chapter on Causal Analysis (Chapter VI).

E. METHODOLOGY

The specification of the model was based on (1) a literature search to identify methodologies and case studies which had been used to determine land uses associated with major projects and (2) the prior experience of personnel with land use planning and forecasting, land use models, impact analyses, and large development projects. In general, the literature search was not productive. However, this does illustrate the potential usefulness of the GEMLUP project.

While many land use models have been developed, these have dealt primarily with total regional growth and activity. They have thus been more complex and at a larger scale than the model we were developing. An example of such a "regional" model is the EMPIRIC activity allocation model, such as that used in the Metropolitan District Commission Wastewater Study [36]. This model allocates projected employment and population to areas within the region based on factors such as vacant land, accessibility, availability of utilities, population characteristics, etc. While useful in showing what influences developed, the model does not address the specific problem we are examining, i.e., induced or associated land uses after a major project has been built.

In addition to "full" models, such as EMPIRIC, there are also many "partial" models. These deal with one section of the economy such as the retail or residential sector. For the most part these models tend to be market and/or location oriented. They are designed to answer either the question of what is the optimum location for that particular type of activity or what is the market for such an activity in a certain geographical area. An example of such a "partial" model is the "Retail Market Potential Model" by Lakshman and Hansen [37]. This model predicts the size and number of retail establishments based on consumer shopping expenditures and accessible to that area. As in the case of the "full" models, these "partial" models are helpful in providing insight on what influences certain land use activities (e.g., relationship between retail establishments and consumer expenditures) they do not, however, establish quantitative relationships between a major project and surrounding land uses.

The literature search also investigated methodologies which have been used in environmental impact studies to predict secondary impacts of major projects. The most useful literature found on this topic was a previous study by the U.S. Environmental Protection Agency titled "Secondary Impacts of Transportation and Wastewater Investments: Review and Bibliography" [38]. This report contained an extensive bibliography and discussion of methodological and studies performed concerning impacts of transportation and wastewater investments. Because our area of interest was residential and office or industrial projects rather than transportation or wastewater investments, many of the impacts and relationships described in this report were not applicable to our project. The report was valuable in helping to direct our thinking. Appendix A contains a bibliography of the sources used in the literature search.

As indicated, the literature search did not produce what we were most interested in finding, viz., quantitative descriptions of land use induced by an association with a major project. Therefore, in developing our causal hypothesis we had to rely primarily on our own past experience and professional judgement, i.e., what we had observed with respect to land use development, particularly when major projects were built.

F. MODEL DESCRIPTION

1. Endogenous Variables

Due to the requirements of the emission factors, the units of the endogenous variables for both the residential model and the industrial/office model are building floor area (except for residential and outdoor recreation land uses and highway lane miles) in each of 12 land use categories. These land case categories are residential, retail, office, manufacturing, whole sale and warehousing, hotel, hospital, cultural, churches, public education, outdoor active recreation, and highway lane miles. These particular categories evolved from a process which balanced the following considerations:

- What land use output was needed for estimating emissions.
- What land use output could most effectively be predicted using a causal model; and
- What land use output would be available during data collection to calibrate the model.

The initial list of the desired model output is shown on Table 3-1. The list was extensive, as the intent was to arrive at the most precise estimate of land use and of emissions as possible. The final list of model output

TABLE 3-1

INITIAL LIST OF ENDOGENOUS VARIABLES

Land Use	Units of Measurement
Residential	
SF attached	Units*
SF detached	Units
2 Family	Units
MF (low rise - 4 stories or less excluding 2 family)	Units Units
MF (high rise - 4 stories or more)	Units Units
Mobile Homes	
Transient Lodgings (Hotels, motels, dorms)	Rooms*
Manufacturing	Total sq.ft.
Transportation services (except airport and warehousing)	Total sq.ft.
Wholesale Trade	Total sq.ft.
Retail Trade (including certain personal services)	Sq.ft. in buildings <25000 ft^2 (<2,323 m^2) 25000-50000 ft^2 (2323-4645 m^2) 50000-100000 ft^2 (4645 - 9290 m^2) >100000 ft^2 (>9290 m^2)

* Will be converted to sq.ft. based on average size per unit or room.

TABLE 3-1 (CONTINUED)

INITIAL LIST OF ENDOGENOUS VARIABLES

Land Use	Units of Measurement
Office (all office including government, utilities and communication offices)	Sq.ft. in buildings <25000 ft^2 (<2,323 m^2) 24000-50000 ft^2 (2323-4645 m^2) 50000-100000 ft^2 (4645-9290 m^2) >100000 ft^2 (>9290 m^2)
Warehousing	Total sq.ft.
Government (non-office)	Total sq.ft.
Cultural	Total sq.ft.
Churches	Total sq.ft.
Hospitals	Sq.ft. in buildings <24000 ft^2 (<2,323 m^2) 24000-50000 ft^2 (2323-4645 m^2) 50000-100000 ft^2 (4645-9290 m^2) >100000 ft^2 (>9290 m^2)
Entertainment Indoors (except sports assembly)	Total sq.ft.
Education, Public	Sq.ft. in buildings. <24000 ft^2 (<2,323 m^2) 24000-50000 ft^2 (2323-4645 m^2) 50000-100000 ft^2 (4645-9290 m^2) >100000 ft^2 (>9290 m^2)
Education (other)	Sq.ft. in buildings <24000 ft^2 (<2,323 m^2) 24000-50000 ft^2 (2323-4645 m^2) 50000-100000 ft^2 (4645-9290 m^2) >100000 ft^2 (>9290 m^2)
Sports Assembly	Total sq.ft.
Airport	Total sq.ft.

TABLE 3-1 (CONTINUED)

INITIAL LIST OF ENDOGENOUS VARIABLES

Land Use	Units of Measurement
Highway right-of-way	Acres
Agriculture	Acres
Mining	Acres
Open Space	Acres

TABLE 3-2

MODEL VARIABLES AND DEFINITIONS

---

1. RES = number of housing units in area of influence in 1970 (excluding major project).

2. COMM = commercial land use in area of influence in 1970 in 1,000 square feet

   Commercial land use includes the following land use codes (LUC) as used by the Public Service Administration Service in its 1962 Land Use Classification Manual.

   LUC 52-59 Retail trade
   61 Personal services
   63 Automobile service
   64 Miscellaneous repair service
   65 Indoor amusement service

3. OFFICE = office land use in area of influence (excluding major project) in 1970 in 1,000 square feet

   LUC 60 Finance, Insurance, Real Estate
   62 Business services
   67 Medical, Health, Legal services
   68 Other professional services

4. MANF = manufacturing land use in area of influence (excluding major project) in 1970 in 1,000 square feet

   LUC 2 Nondurable goods manufacturing
   3 Durable goods manufacturing

5. WHOLE = wholesale/warehouse land use in area of influence in 1970 in 1,000 square feet

   LUC 50 Wholesale
   46 Warehousing

6. HOTEL = Hotel and motel land use in area of influence in 1970 in 1,000 square feet

   LUC 07 Hotels, Motels, Tourist homes

7. HOSPTL = hospital, etc., land use in area of influence in 1970 in 1,000 square feet

   LUC 77 Hospitals, sanatoria, convalescent homes and rest homes

8. CULTUR = cultural land use in area of influence in 1970 in 1,000 square feet

   LUC 76 Museums, libraries, art galleries, except churches (764), arboreta (762), cemeteries (767).

TABLE 3-2 (Continued)

MODEL VARIABLES AND DEFINITIONS

---

9. CHURCH = religious land use in area of influence in 1970 in 1,000 square feet

   LUC 764 Churches
   765 Other religious services

10. EDUC = public educational land use in area of influence in 1970 in 1,000 square feet.

    LUC 74 Public Schools

11. REC = active outdoor recreational land use in area of influence in 1970 in acres

12. (a) HWLMNK = highway land miles in area of influence in 1970, excluding limited access highways.

12. (b) HWLM = highway lane miles in area of influence in 1970

13. (a) MPR70 = residential land use in major project in 1970 in dwelling units

13. (b) MPR68 = residential land use in major project in 1968 in dwelling units

13. (c) MPRt2 = residential land use in major project in base year plus two (t+2) in dwelling units.

14. DUACRE = dwelling units per acre in area of influence in 1960

15. VACACR = percent vacant developable acreage in area of influence in year (t+0)

16. VACHSG = percent vacant housing in area of influence in 1960

17. HWYINT = highway interchanges in area of influence in year (t+5)

18. MINCC = median income of families and individuals in area of influence relative to U.S. median income in 1960

19. INCMP = variable indicating the median income level of major project compared to surrounding community in year (t+2)

---

[1] HWLM without expressways may be used instead, since commercial development may be mostly strip highway development along major arterials

[1] Data for the year (t+2) or 1970 may be used instead

TABLE 3-2 (Continued)
MODEL VARIABLES AND DEFINITIONS

20. OFFVAC = percent office buildings vacant in metropolitan area in year (t+0)

21. OFFACR = office employment per acre in area of influence in year (t+0)

22. DISCBD = distance from center of major project to CBD in year (t+0)

23. ENERGY = cost factor for electricity ($/1500 KWH) for commercial users in the metropolitan area in year (t+0) divided by the average U.S. commercial rate in 1960

24. RRMI = railroad mileage in area of influence in year (t+0)

25. WWEA = warehouse and wholesale employment per acre in area of influence in year (t+0)

26. EMPACR = total employment per acre in area of influence in year (t+0)

27. NONHSE = nonhousehold population per acre in area of influence in 1960

28. MPKIDS = schoolage children per dwelling unit in major project in year (t+2)

29. ENRACR = public school enrollment per acre in area of influence in 1960

30. MANACR = manufacturing employment per acre in area of influence in year (t+0)

31. DELPOP = growth factor for total regional population between 1960 and 1970 (county data)

32. DELEMP = growth factor for total regional employment between 1960 and 1970 (county data)

33. MINCR = median income of the region in year (t+0) relative to the median U.S. income in 1960

34. (a) MPE70 = number of employees in major project in 1970

34. (b) MPE68 = number of employees in major project in 1968

34. (c) MPEt2 = number of employees in major project in base year (t+2)

35. AUTO = automobile drivers per acre in county in 1960

(the first 12 variables on Table (3-2) was a much simplified version of the earlier list.) This simplified version resulted from these three decisions:

- It was decided to aggregate categories of land use (i.e., residential units, total office square feet, etc.) rather than use the more specific subcategories within these broad categories (i.e., number of two family units or office square feet in buildings greater than 100,000 square feet, 9.3×10^3 square meters). We made the decision as we felt our causal hypothesis would be more accurate and thus more successfully tested at this larger scale. Further disaggregation would be performed in the model calibration phase, after the aggregated model had been tested.
- Certain of the initial land use categories were combined based on the availability of data and our interest in simplifying the causal structure. For example, warehousing and wholesale trade were combined. In each of these cases, it was felt that the disaggregated categories offered no further accuracy in predicting emissions than the aggregate category.
- Certain of the land use categories were excluded from the model as either the categories were independent of the major project (i.e., correctional institutions) or the categories produced only insignificant quantities of emissions. The categories eliminated were: government (non-office) such as correctional institutions and military installations; airports; sports assemblies; educational uses except public education; passive open space; agricultural and mining.

2. Exogenous Variables

The model consists of 23 independent variables. These variables represent (the numbers refer to the order on Table 3-2):

- Population housing and employment characteristics (variables 13, 14, 16, 18, 19, 20, 21, 25, 26, 27, 28, 29, 30, 34)
- Accessibility measures (variables 17, 22, 24)
- Developability measure (variable 15)
- Regional influences (variables 23, 31, 32, 33, 35)

The independent variables for each equation were selected because of their perceived causal relationship with the dependent variables. Prior to selecting these variables a complete list of all possible factors influencing the dependent variable was prepared. From this list the most significant variables were identified.

The specific format for each independent variable was developed based on 1) availability of data; 2) consistency of data among case studies; and 3) the appropriate time period for the data. Several of the major decisions made in this regard were:

a. Availability of Data

For many of the independent variables (variables 14, 15, 16, 17, 18, 21, 24, 25, 26, 27, 29, and 30) the desired level of aggregation for the data item was by area of influence (i.e., within the 10,000 acres). Because the area of influence of the project did not coincide with boundaries of recognizable geographical divisions, such as municipal boundaries or census tract boundaries, obtaining information for certain independent variables for area of influence was often very difficult. In these cases the information was collected for the census tracts comprising the area of influence. Thus, the level of aggregation for variables 14, 16, 18, 27, and 29 was by census tracts; for variables 15, 17, 21, 24, 25, 26, 30 by area of influence. The distinction between the two groups was how the information was most easily available.

The level of aggregation for certain variables was regional. In these cases the variable was expressed at that level either because the regional level was the smallest unit where accurate information was readily available (variables 20 and 23) or because the variable was expressing regional influences on the land uses within the area of influence (variables 31, 32, 33 and 35).

b. Consistency of Data

Many of the independent variables (variables 14, 15, 16, 20, 21, 23, 25, 26, 27, 28, 29, 30, 31, 32, and 33) were expressed as density factors or as percentages rather than in absolute numbers. The intent was to assure that the data was consistent among case studies.

(1) Time Periods. The time period for the majority of independent variables is either the base year (t+0) when the project started or 1960. The reason some variables are for the period t+0 and others for 1960 is due to data constraints.

The basic objective was to have all these variables expressed for the base year (i.e., baseline conditions prior to the start of the project). Where it was very difficult to obtain data for the base year, then census data for the year 1960 was used. The specific data items for 1960 are variables 14, 16, 18, 27, 29, 33. Those for the base year are variables 20, 21, 22, 23, 24, 25, 26, 30, and 35.

The independent variables describing the major characteristics of the project are, for the time period t+2 (i.e., variables 19 and 28). The independent variables measuring the size of the major project (13a, 13b, 13c, and 34a, 34b, 34c) consist of observations of major project size in three time periods, two years after initial occupancy (t+2), 1968, and 1970. While it was originally planned to collect only the final size of the major project (i.e., 1970), the possible differences in the phasing of occupancy of the major project could lead to differing amounts of induced land use. Consequently, it was decided to specify three independent variables measuring the size of the major project.

Variables 31 and 32 were designed to measure the influence of regional growth on the local land uses. The period between 1960 and 1970 was selected as census data was readily available to determine growth and the period coincided fairly well with the initial year of operation of the projects in our case studies and 1970 - the year in which land use impacts were assumed to have stabilized.

The only variable not available for the period t+0 1960, t+2, 1968, or 1970 was variable 17, numbers of highway interchanges. The time period for this variable was t+5. The reason was that the building of new highway interchanges is commonly known five years in advance.

3. Simultaneous Block

The specification of the model resulted in twelve structural equations predicting the twelve endogenous variables described above. In both models, five of these equations are simultaneous. These five equations are shown on Tables 3-3 and 3-4. The path diagrams representing these equations are shown in Figures 3-4 and 3-5.

4. Recursive Block

The remaining seven equations in each model are recursive. These equations are shown on Table 3-5 and Table 3-6; their path diagrams are shown in Figures 3-6 and 3-7. Finally the complete model is shown on Tables 3-7 and 3-8 and in Figures 3-8 and 3-9.

G. THEORETICAL AND EMPIRICAL BASIS FOR THE MODEL

This section describes the theoretical or empirical base for the residential project model and for the industrial/office project model. Because the equations for these models are in many cases the same and in other cases very similar, the two models are discussed jointly. The three major project variables are treated as one variable (MPR or MPE) in the following discussion.

1. Equation 1

Res. Proj. RES = OFFICE + MANF + HWLM +
DUACRE + VACACR +
VACHSG + DELPOP + MPR

Ind/Off. Proj. RES = OFFICE + MANF + HWLM +
DUACRE + VACACR - VACHSG +
DELPOP + MPE

In these equations residential development (1970) is a function of office development (1970); manufacturing development (1970); highway lane miles (1970); dwelling units per acre (1960); percent vacant developable

TABLE 3-3

FIVE EQUATION SIMULTANEOUS BLOCK IN THE RESIDENTIAL MODEL

RES = OFFICE + MANF + HWLM + MPR + DUACRE + VACACR + VACHSG + DELPOP

COMM = RES + OFFICE + MANF + HWLM + MPR70 + HWYINT + MINCC + INCMP + DELPOP + MINCR

OFFICE= RES + COMM + MANF + HWLM + MPR70 + VACACR + HWYINT + OFFVAC + OFFACR + DISCBD + DELEMP

MANF = HWLM + MPR70 + HWYINT + ENERGY + RRMI + MANACR + DELEMP

HWLM = RES + COMM + OFFICE + MANF + MPR70 + HWYINT + DISCBD + EMPACR

TABLE 3-4

FIVE EQUATION SIMULTANEOUS BLOCK IN THE INDUSTRIAL-OFFICE MODEL

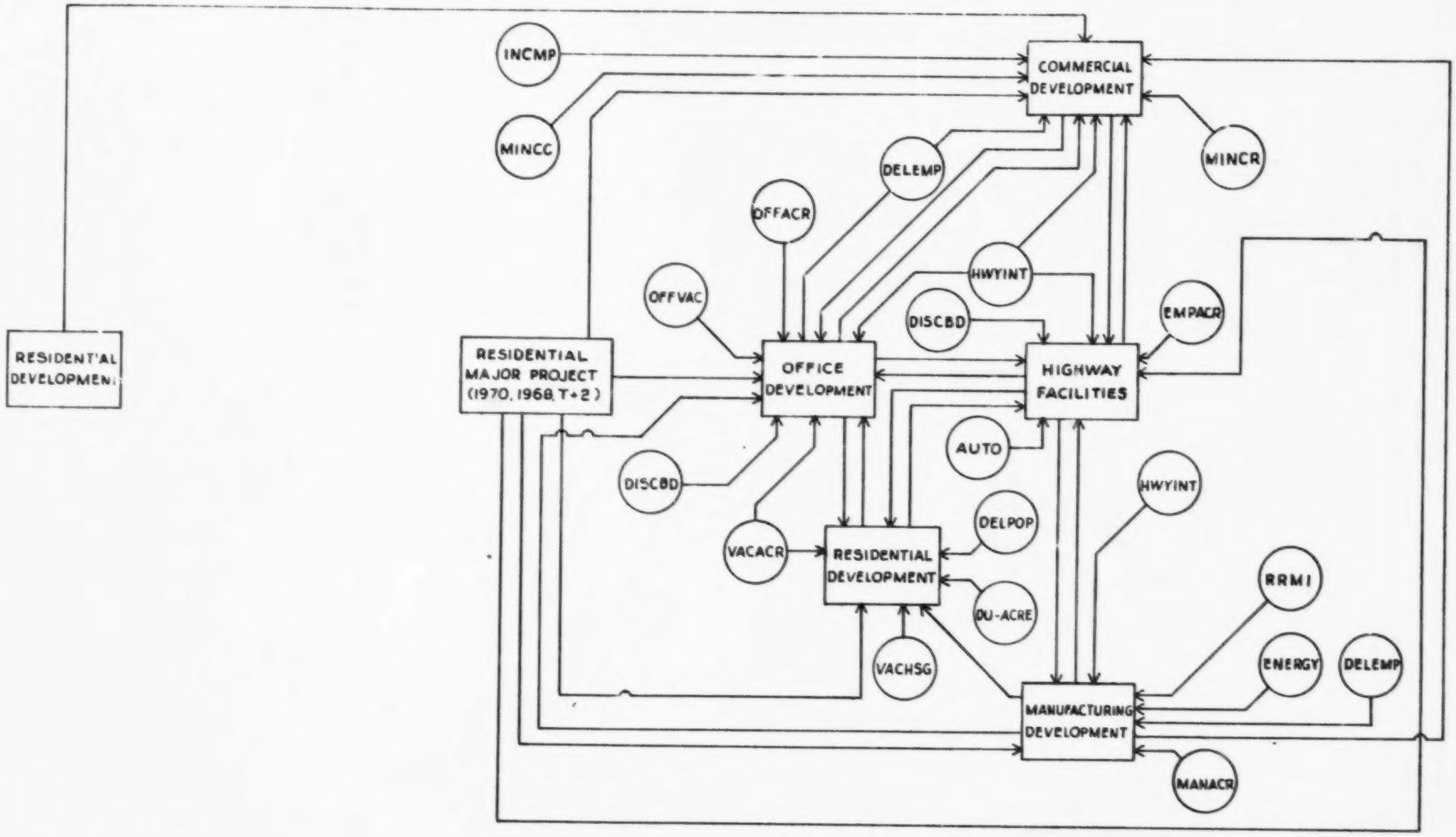
RES = OFFICE + MANF + HWLM + MPE + DUACRE + VACACR + VACHSG + DELPOP

COMM = RES + OFFICE + MANF + HWLM + MPE70 + HWYINT + MINCC + DELEMP + INCMP

OFFICE= RES + RETAIL + MANF + HWLM + MPE70 + VACACR + HWYINT + OFFVAC + OFFACR + DISCBD + EMPACR

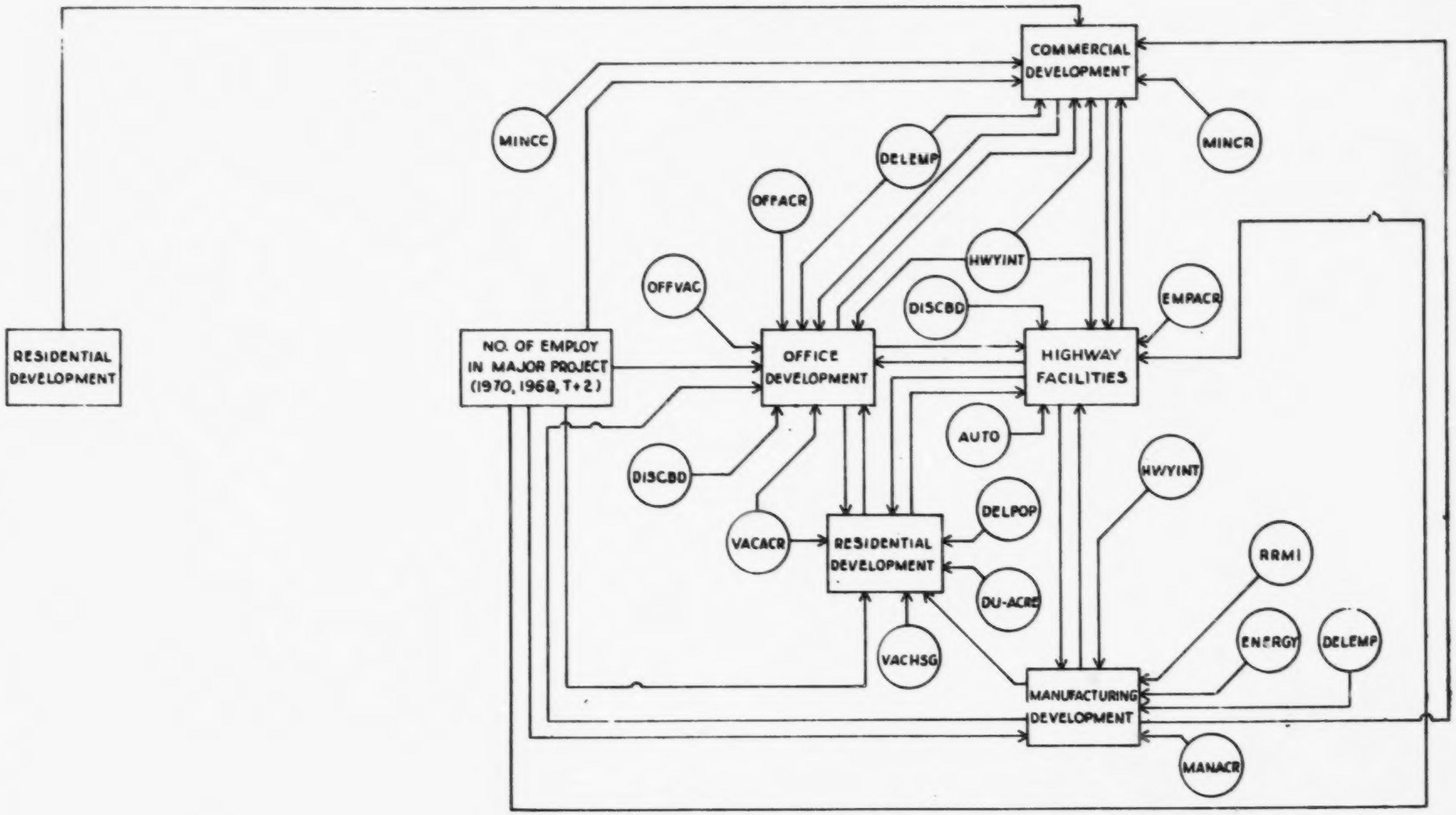
MANF = HWLM + MPE + HWYINT + ENERGY + RRMI + MANACR + DELEMP

HWLM = RES + RETAIL + OFFICE + MANF + MPE70 + HWYINT + DISCBD + EMPACR


RESIDENT'AL DEVELOPMEN
INCMP
MINCC
COMMERCIAL DEVELOPMENT
MINCR
DELEMP
OFFACR
OFFVAC
HWYINT
DISCBD
EMPACR
RESIDENTIAL MAJOR PROJECT (1970, 1968, T+2)
OFFICE DEVELOPMENT
HIGHWAY FACILITIES
AUTO
DISCBD
HWYINT
DELPOP
VACACR
RESIDENTIAL DEVELOPMENT
DU-ACRE
RRMI
VACHSG
ENERGY
DELEMP
MANUFACTURING DEVELOPMENT
MANACR

NOTE: SOME OF THE VARIABLES APPEAR TWICE (E.G., RESIDENTIAL) FOR EASE OF PRESENTATION

Figure 3-4 Five Equation Simultaneous Block In The Residential Model



RESIDENTIAL DEVELOPMENT
MINCC
COMMERCIAL DEVELOPMENT
MINCR
DELEMP
OFFACR
HWYINT
OFFVAC
DISCBD
EMPACR
NO. OF EMPLOY IN MAJOR PROJECT (1970, 1968, T+2)
OFFICE DEVELOPMENT
HIGHWAY FACILITIES
AUTO
DISCBD
HWYINT
VACACR
DELPOP
RESIDENTIAL DEVELOPMENT
DU-ACRE
RRMI
VACHSG
ENERGY
DELEMP
MANUFACTURING DEVELOPMENT
MANACR

NOTE: SOME OF THE VARIABLES APPEAR TWICE (E.G., RESIDENTIAL) FOR EASE OF PRESENTATION

Figure 3-5 Five Equation Simultaneous Block In The Industrial-Office Model

TABLE 3-5

SEVEN EQUATION RECURSIVE BLOCK IN THE RESIDENTIAL MODEL

WHOLE = COMM + MANF + HWLM + HWYINT + RRMI + WWEA + DELEMP

HOTEL = RES + OFFICE + MANF + MPR + DISCBD + EMPACR

HOSPTL = RES + MPR + DISCBD + NONHSE

CULTUR = RES + MPR + DISCBD

CHURCH = RES + MPR + MPKIDS

EDUC = RES + MPR + MPKIDS + ENRACR

REC = RES + MPR + MINCC + INCMP

TABLE 3-6

SEVEN EQUATION RECURSIVE BLOCK IN THE INDUSTRIAL/OFFICE MODEL

WHOLE = RETAIL + MANF + HWLM + HWYINT + RRMI + WWEA + DELEMP

HOTEL = RES + OFFICE + MANF + MPE + DISCBD + EMPACR

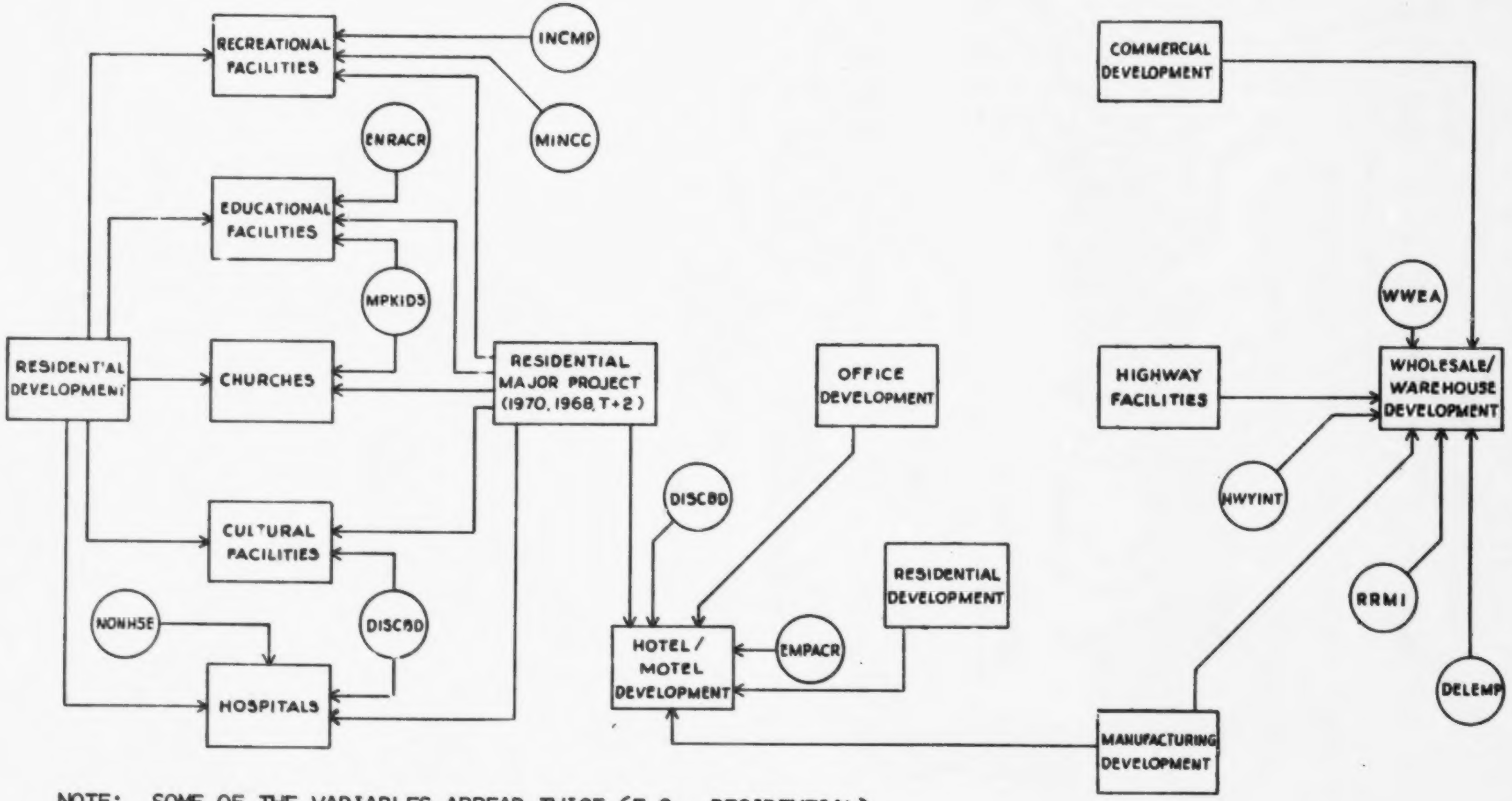
HOSPTL = RES + MPE + DISCBD + NONHSE

CULTUR = RES + DISCBD

CHURCH = RES

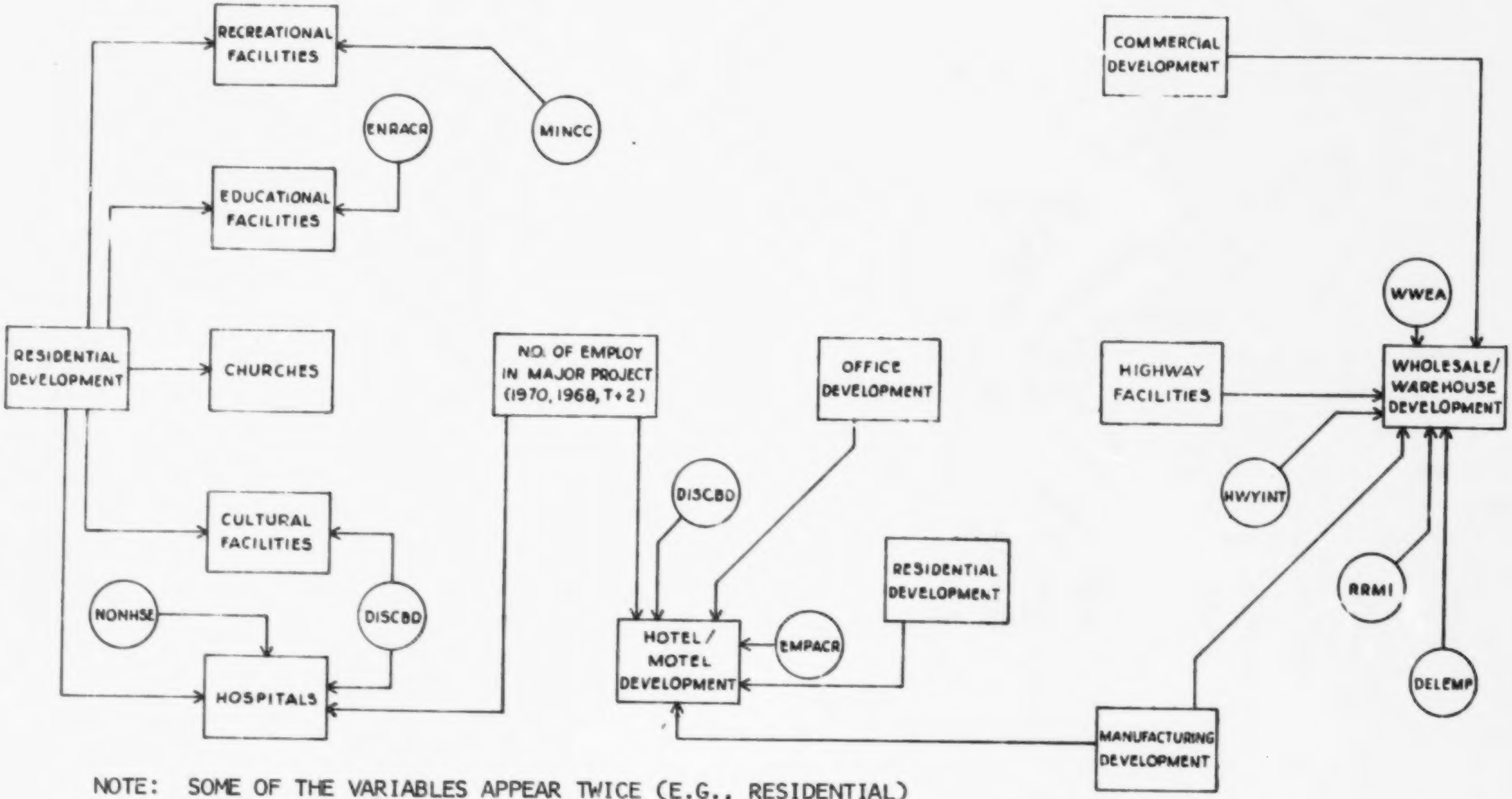
EDUC = ENRACR

REC = RES + MINCC


RECREATIONAL FACILITIES
INCMP
MINCC
ENRACR
EDUCATIONAL FACILITIES
MPKIDS
RESIDENT'AL DEVELOPMENT
CHURCHES
RESIDENTIAL MAJOR PROJECT (1970, 1968, T+2)
CULTURAL FACILITIES
DISCBD
NONHSE
HOSPITALS
OFFICE DEVELOPMENT
DISCBD
RESIDENTIAL DEVELOPMENT
EMPACR
HOTEL / MOTEL DEVELOPMENT
COMMERCIAL DEVELOPMENT
HIGHWAY FACILITIES
WWEA
WHOLESALE/ WAREHOUSE DEVELOPMENT
HWYINT
RRMI
DELEMP
MANUFACTURING DEVELOPMENT

NOTE: SOME OF THE VARIABLES APPEAR TWICE (E.G., RESIDENTIAL) FOR EASE OF PRESENTATION

Figure 3-6 Seven Equation Recursive Block In The Residential Model


RECREATIONAL FACILITIES
EDUCATIONAL FACILITIES
CHURCHES
CULTURAL FACILITIES
HOSPITALS
RESIDENTIAL DEVELOPMENT
MINCC
ENRACR
DISCBD
NONHSE
NO. OF EMPLOY IN MAJOR PROJECT (1970, 1968, T+2)
DISCBD
OFFICE DEVELOPMENT
HOTEL / MOTEL DEVELOPMENT
EMPACR
RESIDENTIAL DEVELOPMENT
COMMERCIAL DEVELOPMENT
HIGHWAY FACILITIES
WWEA
WHOLESALE/ WAREHOUSE DEVELOPMENT
HWYINT
RRMI
DELEMP
MANUFACTURING DEVELOPMENT

NOTE: SOME OF THE VARIABLES APPEAR TWICE (E.G., RESIDENTIAL) FOR EASE OF PRESENTATION

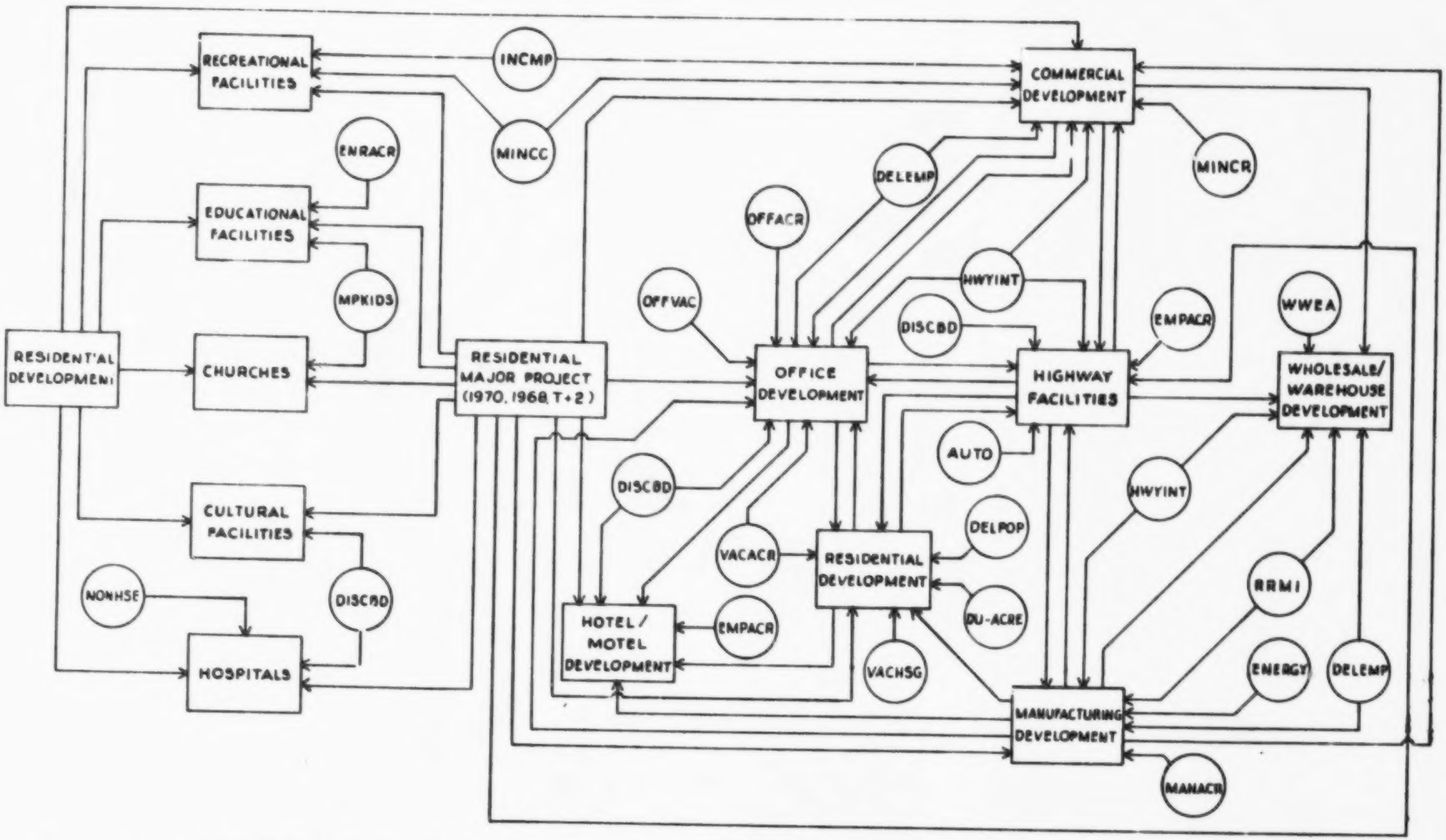
Figure 3-7 Seven Equation Recursive Block In The Industrial-Office Model

Table 3-7 Original Specification Of The Residential Model

	RES	COMM	OFFICE	MANF	HWLM	No. of Housing Units, Major Project	D.U./Acre, Census Tract, Base Year	Percent Vacant Land, Area of Influence, Base Year	Housing Vacancy, Census Tract, 1960	No. of Highway Interchanges, Area of Influence, Base Year + 5	Median Income, Census Tract, 1960	Income Level, Base Year + 2	Office Vacancy, Region, Base Year	Office Emp/Acre, Municipality, Base Year	Distance to CBD, Base Year	Energy Cost, 1500 kWh	Miles of Railroad, Area of Influence, Base Year	Warehouse-Wholesale, Emp/Acre, Base Year	Total Emp/Acre, Municipality, Base Year	Non-Household, Population/Acre, Census Tract	H.P. Children/Unit, Base Year + 2	Enrollment Acre, Municipality, Base Year	Manf Emp/Acre, Municipality, Base Year	Ten Years, Regional Growth		Median Income, Region, Base Year	Auto Drivers Per Acre
																								ΔEMP	ΔPOP		
1. RES =			+ OFFICE	+ MANF	+ HWLM	+ MPR	+ DUACRE	+ VACACR	+ VACHSG																		
2. COMM =	RES		+ OFFICE	+ MA'F	+ HWLM	+ MPR				+ HWYINT	+ MINCC	+ INC													+ DELPOP		
3. OFFICE =	RES	+ COMM		+ MANF	+ HWLM	+ MPR		+ VACACR		+ HWYINT			+ OFFVAC	+ OFFACR	+ DISCBD									+ DELEMP		+ MINCR	
4. MAUF =					+ HWLM	+ MPR				+ HWYINT						+ ENERGY	+ RRMI							+ DELEMP			
5. WHOLE =		+ COMM		+ MANF	+ HWLM					+ HWYINT							+ RRMI	+ WHEA					+ MANACR	+ DELEMP			
6. HOTEL =	RES		+ OFFICE	+ MANF		+ MPR									+ DISCBD				+ EMPACR					+ DELEMP			
7. HOSPTL =	RES					+ MPR									+ DISCBD					+ NONHSE							
8. CULTUR =	RES					+ MPR									+ DISCBD												
9. CHURCH =	RES					+ MPR																					
10. EDUC =	RES					+ MPR																					
11. REC =	RES					+ MPR					+ MINCC	+ INC															
12. HWLM =	RES	+ COMM	+ OFFICE	+ MANF		+ MPR				+ HWYINT					+ DISCBD				+ EMPACR								+ AUTO.

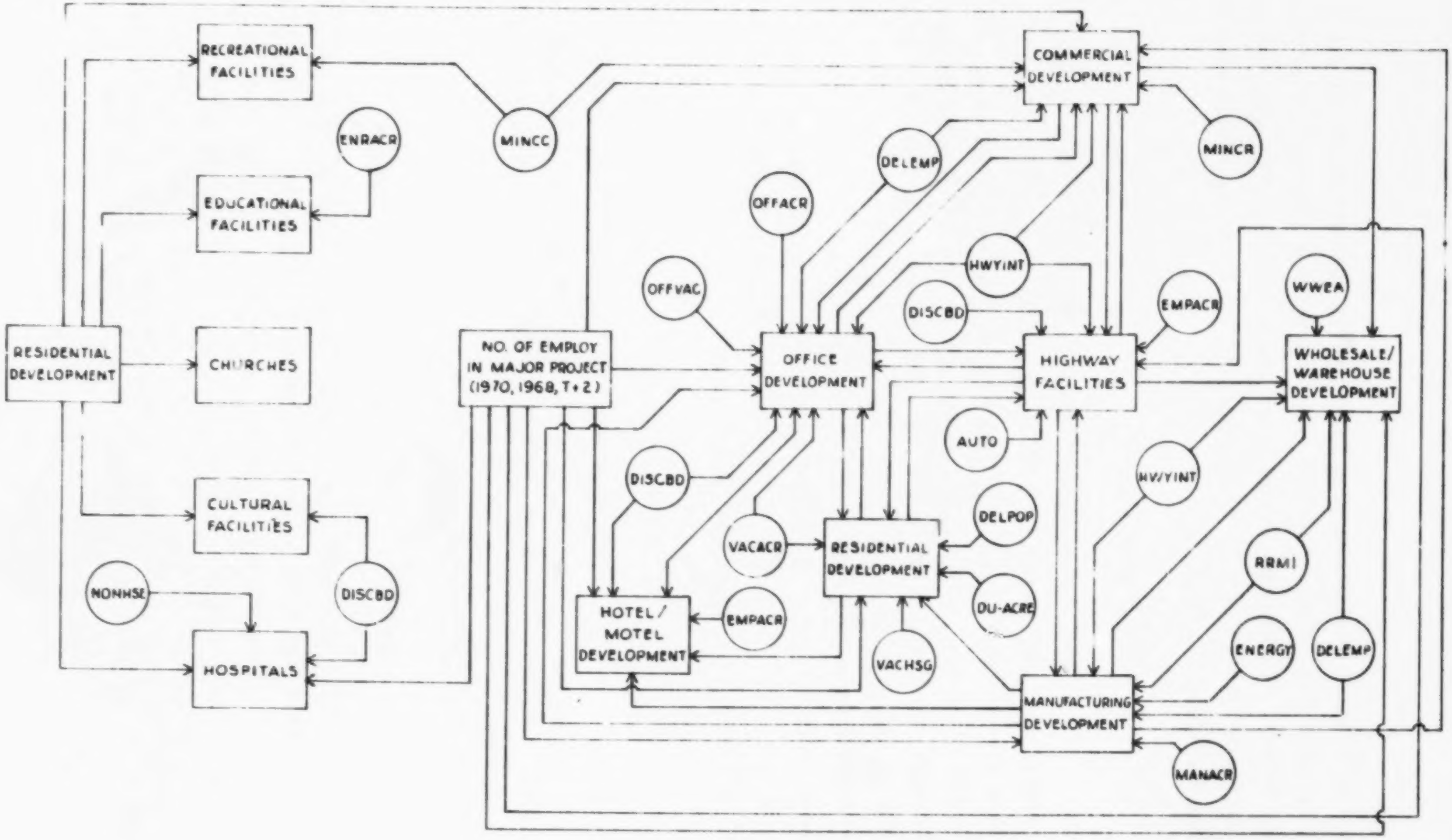
Table 3-8 Original Specification Of The Industrial-Office Model

	RES	COMM (RETAIL)	OFFICE	MANF	HWLM	No. of Employees, Major Project	D.U./Acre, Census Tract, Base Year	Percent Vacant Land, Area of Influence, Base Year	Housing Vacancy, Census Tract, 1960	No. of Highway Interchanges, Area of Influence, Base Year + 5	Median Income, Census Tracts, 1960	Office Vacancy, Region, Base Year	Office Employee/Acre, Municipality, Base Year	Distance to CBD, Base Year	Energy Cost, 1500 kWh	Miles Railroad, Area of Influence, Base Year	Warehouse-Wholesale, Employee/Acre, Base Year	Total Employee/Acre, Municipality, Base Year	Non-Household, Population/Acre, Census Tract, 1960	M.P. Children/Unit, Base Year + 2	Enrollment/Acre, Municipality, Base Year	Manf Employee/Acre, Municipality, Base Year	Ten Years, Regional Growth		Median Income, Region, Base Year	Auto Drivers, Per Acre
																							ΔEMP	ΔPOP		
1. RES =			+ OFFICE	+ MANF	+ HWLM	+ MPE	+ DUACRE	+ VACACR	+ VACHSG															+ DELPOP		
2. COMM =	RES		+ OFFICE	+ MANF	+ HWLM	+ MPE				+ HWYINT	+ MINCC												+ DELEMP		+ MINCR	
3. OFFICE =	RES	+ COMM		+ MANF	+ HWLM	+ MPE		+ VACACR		+ HWYINT		+ OFFVAC	+ OFFACR	+ DISCBD									+ DELEMP			
4. MANF =					+ HWLM	+ MPE				+ HWYINT					+ ENERGY	+ RRMI						+ MANACR	+ DELEMP			
5. WHOLE =		+ COMM		+ MANF	+ HWLM					+ HWYINT						+ RRMI	+ WHEA						+ DELEMP			
6. HOTEL =	RES		+ OFFICE	+ MANF		+ MPE								+ DISCBD				+ EMPACR								
7. HOSPTL =	RES					+ MPE								+ DISCBD					+ NONHSE							
8. CULTUR =	RES													+ DISCBD												
9. CHURCH =	RES																			+ MPKIDS						
10. ED =	RES																			+ MPKIDS	+ ENRACR					
11. REC =	RES										+ MINCC															
12. HWLM =	RES	+ COMM	+ OFFICE	+ MANF		+ MPE				+ HWYINT				+ DISCBD				+ EMPACR								+ AUTO


RECREATIONAL FACILITIES
INCMP
COMMERCIAL DEVELOPMENT
MINCC
ENRACR
EDUCATIONAL FACILITIES
DELEMP
OFFACR
MINCR
HWYINT
MPKIDS
OFFVAC
DISCBD
EMPACR
WWEA
RESIDENT'AL DEVELOPMENI
CHURCHES
RESIDENTIAL MAJOR PROJECT (1970, 1968, T+2)
OFFICE DEVELOPMENT
HIGHWAY FACILITIES
WHOLESALE/ WAREHOUSE DEVELOPMENT
AUTO
DISCBD
HWYINT
CULTURAL FACILITIES
DELPOP
VACACR
RESIDENTIAL DEVELOPMENT
NONHSE
DISCBD
RRMI
DU-ACRE
HOTEL/ MOTEL DEVELOPMENT
EMPACR
HOSPITALS
VACHSG
ENERGY
DELEMP
MANUFACTURING DEVELOPMENT
MANACR

NOTE: SOME OF THE VARIABLES APPEAR TWICE (E.G., RESIDENTIAL) FOR EASE OF PRESENTATION

Figure 3-8 Original Specification Of The Residential Model


RECREATIONAL FACILITIES
EDUCATIONAL FACILITIES
CHURCHES
CULTURAL FACILITIES
HOSPITALS
RESIDENTIAL DEVELOPMENT
ENRACR
MINCC
DISCBD
NONHSE
NO. OF EMPLOY IN MAJOR PROJECT (1970, 1968, T+2)
COMMERCIAL DEVELOPMENT
MINCR
DELEMP
OFFACR
OFFVAC
HWYINT
DISCBD
EMPACR
WWEA
OFFICE DEVELOPMENT
HIGHWAY FACILITIES
WHOLESALE/ WAREHOUSE DEVELOPMENT
AUTO
DISCBD
HW/YINT
VACACR
RESIDENTIAL DEVELOPMENT
DELPOP
DU-ACRE
RRMI
HOTEL/ MOTEL DEVELOPMENT
EMPACR
VACHSG
ENERGY
DELEMP
MANUFACTURING DEVELOPMENT
MANACR

NOTE: SOME OF THE VARIABLES APPEAR TWICE (E.G., RESIDENTIAL) FOR EASE OF PRESENTATION

Figure 3-9 Original Specification Of The Industrial-Office Model

land (base year); housing vacancy rate (1960); and the number of dwelling units in the major project if the project is a residential project; or the number of employees in the major project if the project is an industrial/office project. The reasons each of these variables were included are as follows:

a. OFFICE and MANF

Office and manufacturing development (1970) may both provide significant employment opportunities. They thus create a demand and pressure for residential housing located near the employment base. This relationship between residential housing and office and manufacturing development was shown in the use of the EMPIRIC Model by Metcalf & Eddy for the MDC Wastewater Study, [36]. The model identified a significant correlation between the change in manufacturing and white collar employment and households. Therefore, we can expect that if all other variables in the equation are kept constant that residential housing will increase as the office employment or manufacturing employment increases.*

b. HWLM

Highway lane miles were included as a causal influence on residential development as such development (particularly higher density development) tends to locate in areas well served by transportation networks. Use of the EMPIRIC Model identified a correlation between highway miles and residential development.

c. DUACRE

It is expected that if all other variables in the equation are held constant, the total amount of residential development in 1970 will be influenced by the residential density in the area in 1960 The greater the density in 1960, the greater the residential development in 1970.

---

* The assumption that all variables in the equation are kept constant except the specific variable described applies throughout this section.

d. VACACR

A prime factor encouraging new residential development is availability of vacant developable land. Therefore, in our model we can expect a causal relationship between vacant developable land (1960) and total residential development (1970).

e. VACHSG

Vacancy rates have traditionally been used in planning studies as indicators of the demand per housing. Normal rates necessary to provide an adequate measure of freedom of choice and housing opportunity are 2 percent for owner-occupied units and 6 percent for renter-occupied units. Low vacancy rates indicate a high demand for additional housing, while high vacancy rates indicate a low demand. The theory in our model is that the housing vacancy rate in 1960 is a negative causal factor on the level of residential development in 1970.

f. DELPOP

Regional population growth (1960-1970) was included in the model to control for the influence of regional growth on local residential development. The theory is that where regional population growth is large there will be increased demand for residential development in the local area of influence.

g. MPR

In the residential project model, the number of units of the project was included as a causal influence on residential development. The theory is that the major project will stimulate other residential housing in the area by increasing the attractiveness of the area to developers.*

---

* It is possible that at a certain size the residential project may capture the demand for new housing in the area and thus have a negative impact on other new residential development.

h. MPE

In the industrial/office project model the number of employees of the project was included as a causal influence on residential development. The theory is that the employees will contribute to the demand for housing in the area and thus positively affect residential development.

2. Equation 2

Res. Proj. COMM = RES + OFFICE + MANF + HWLM + HWYINT + MINCC + INCMP + DELEMP + MINCR + MPR

Ind/Off. Proj. COMM = RES + OFFICE + MANF + HWLM + HWYINT + MINCC + DELEMP + MINCR + MPE

In these equations commercial development (1970) is a function of residential uses (1970), office uses (1970), manufacturing uses (1970), highway lane miles (1970), number of highway interchanges (base year + five), median income (1960), regional employment growth (1960-1970), median regional income (base year), and area in the case of the residential project, the number of units and median income (base year + two) of the major project and, in the case of the industrial/office project, the number of employees of the major project. The reasons each of these variables were included are as follows:

a. RES, OFFICE, MANF

Residential, office, and manufacturing development were included as each of those stimulates a demand for commercial services. Thus the greater each of these uses are, the more the commercial development. Use of the EMPIRIC Model identified a significant correlation between the growth of residential development and blue collar employment.*

---

* In the EMPIRIC Model, commercial development included both retail trade and office uses.

b. HWLM

In numerous land use studies, it is commonly observed that commercial activity tends to locate in areas with good transportation access. Therefore, highway lane miles, as an indication of transportation access, was chosen as a causal factor influencing commercial development.

c. HWYINT

Number of highway interchanges is a measure of an area's accessibility. Land use studies have identified limited access highway interchanges as a strong influence in the location of commercial development. Therefore, we can expect a causal relationship between the number of highway interchanges and the amount of commercial development.

d. MINCC

Expenditures on commercial services are a function of income. Thus the higher the income within the area of influence the greater the amount of commercial services that can be supported.

e. DELEMP

Commercial development both serves and is a part of the region's employment base. Thus we can expect that growth in regional employment (1960-1970) will have a positive impact on commercial development within the area of influence.

f. MINCR

As indicated on (d) above, expenditures on commercial services are a function of income. Thus the higher the regional income, the greater the expenditures on commercial development within the local area of influence provided all other variables in the equation are held constant.

g. INCMP + MPR

The number of units of the residential project and the income level of residents in the project will both add to the demand for commercial services in the area of influence. Thus both variables were included in the residential model as causal influences of commercial development.

h. MPE

The number of employees of the industrial/office project will contribute to the demand for commercial services within the area of influence. We can expect, therefore, that the greater the number of employees, the greater the amount of commercial services.

3. Equation 3

Res. Proj. OFFICE = RES + COMM + MANF +
HWLM + VACACR +
HWYINT - OFFVAC + OFFACR -
DISCBD + DELEMP + MPR

Ind/Off. Proj. OFFICE = RES + COMM +
MANF + HWLM + VACACR + HWYINT +
OFFVAC - OFFACR - DISCBD +
DELEMP + MPE

In these equations office development is a function of residential development (1970), commercial development (1970), manufacturing development (1970), highway lane miles (1970), percent vacant land (base year), number of highway interchanges (base year + five), regional office vacancy rate (base year), office employees/acre (base year), distance from the central business district (base year), regional employment growth (1960-1970); and, in the case of the residential project, the number of units of the major project and, in the case of the industrial/office project, the number of employees (1970) of the major project.

The reasons each of these variables were included are as follows:

a. RES

Residential development stimulates office development by creating a market for office services. As indicated, use of the EMPIRIC Model identified a significant correlation between the growth of commercial (office) development and residential growth.

b. COMM, MANF

A survey by the Fairfax County, Virginia Industrial Authority in 1970 [43] identified proximity to commercial and industrial development as strong factors in office building location. The commercial development provides services to the office building employees, while the manufacturing development creates demand for the services of people working in the office building. Thus both types of development were included in our model as causal influences of office development.

c. HWLM

Highway lane miles (1970) are included as office development tends to locate in areas which are well served by transportation systems.

d. VACACR

New office development locates in areas where vacant developable land is available. This causal relationship was identified in land use studies using the EMPIRIC Model. Thus percent vacant developable land was included as a causal influence or office development. Where land is available, the office development will be greater than in areas where land is more limited.

e. HWYINT

Office development tends to locate in areas which are easily accessible to population and employment centers. Therefore, we can expect a causal relationship between the number of highway interchanges in an area and the amount of office development that occurs.

f. OFFVAC

Regional office vacancy rate (base year) was included in the equation; the theory is that office development (1970) is inversely related to the regional office vacancy rate (base year). The lower the vacancy rate, the greater the demand for new office development and the greater the amount of office development in the area of influence in 1970.

g. OFFACR

It is expected that holding all other variables in the equation constant, the total amount of office development in 1970 will be influenced by the density of office employees in the area of 1960. The greater the density in 1960, the greater the office development in 1970.

h. DISCBD

As indicated, office development tends to locate in areas accessible to population and employment centers. Thus we can expect that as the distance to the central business district increases, the amount of office development decreases. Use of the EMPIRIC Model identified access to total households within 10 miles to be a significant factor in the location of office development.

i. DELEMP

Office development is part of the region's employment base. It can be expected, therefore, that growth in the region's employment (1960-1970) will have a positive impact on office development within the area of influence.

j. MPR

In the residential project model the number of units of the residential model was included as a causal influence of office development. The theory is that the major project will create additional demand for office services.

k. MPE

In the industrial/office model the number of employees of the major project was included as a causal influence of office development. The theory is that the major project will attract associated office development.

4. Equation 4

Res. Proj. MANF = HWLM +
HWYINT - ENERGY + RRMI +
MANACR + DELEMP + MPR

Ind/Off. Proj. MANF = HWLM + HWYINT -
ENERGY + RRMI + MANACR +
DELEMP + MPE

In these equations manufacturing development (1970) is a function of highway lane miles (1970); number of highway interchanges (base year +5); the cost of energy (base year); railroad mileage in the area of influence (base year); regional employment growth (1960-1970); and in the case of the residential project, the number of units of the major project and, in the case of the office/industrial project, the number of employees of the major project. The reasons each of these variables were included are as follows:

a. HWLM, HWYINT, RRMI

Prime location requirements for manufacturing industries include proximity to good highways and availability of railroad access. Therefore, highway lane miles, number of highway interchanges and railroad mileage were all included as causal influences of manufacturing development on the theory that as each of these variables increases, the desirability of the area for manufacturing development increases and, therefore, the amount of manufacturing development increases.

b. ENERGY

The MDC Wastewater Study, 1973 [36] identified the cost of energy as an influence on the location of manufacturing industries. The theory in the model is that the lower the cost of energy the more attractive the area is for manufacturers.

c. MANACR

It is expected that holding all other variables in the equation constant the total amount of manufacturing development in 1970 will be influenced by the density of manufacturing employees in 1960. The greater the density in 1960, the greater the manufacturing uses in 1970.

d. DELEMP

Manufacturing employment is part of the region's employment base. We can expect, therefore, that an increase in the region's employment will have a positive impact on manufacturing uses in the area of influence.*

e. MPR

For the residential project, the number of units of the project should have a positive influence on manufacturing development in the area as the project provides a new labor supply, thus encouraging firms to locate in the area.

f. MPE

For the industrial/office project, the number of employees of the project should have a positive impact on manufacturing uses in the area. Associated firms tend to be attracted to areas where a major manufacturing use is located.

---

* This impact may not be felt if the region's employment as a whole increased, but the manufacturing component declines.

5. Equation 5

WHOLE = COMM + MANF + HWLM + HWYINT +
RRMI + WWEA + DELEMP

In this equation wholesale/warehouse development (1970) is a function of commercial uses (1970); manufacturing uses (1970); highway lane miles (1970); number of highway interchanges (base year +5); railroad mileage in the area of influence (base year); and the density of wholesale/warehousing employees (base year) and regional employment growth. The reasons each of these variables were included are as follows:

a. COMM + MANF

Land use studies have shown that commercial and manufacturing uses tend to attract associated warehousing and wholesale uses. Therefore, commercial and manufacturing uses were included in our model as causal influences of wholesale/warehousing uses.

b. HWLM, HWYINT, RRMI

Locational factors which attract wholesale/warehouse development are accessibility to highways and availability of railroads. Therefore, highway lane miles, number of highway interchanges, and railroad mileage were all included in the model as causal influences of wholesale/warehouse development on the theory that as each of these variables increases the desirability of the area for wholesale/warehousing development increases and, therefore, the amount of wholesale/warehouse development increases.

c. WWEA

It is expected that holding all other variables in the equation constant, the total amount of wholesale/warehouse development in 1970 will be influenced by the density of wholesale/warehouse employees in 1960. The greater the density in 1960, the greater the wholesale/warehouse employees in 1970.

d. DELEMP

Wholesale/warehouse development part of the region's employment base is also influenced by the commercial and manufacturing sector of this base. We can expect, therefore, that as the region's employment base increases wholesale/warehouses development in the area of influence is positively affected.

It should be noted that in these equations the major project is not included as a causal influence for warehousing/wholesale development. In the case of the residential project it was felt that there was not a relationship between the project and warehouse/wholesale development. For the industrial/office project there may be some relationship, but we felt that it was less significant than the other relationships identified.

6. Equation 6

Res. Proj. HOTEL = RES + OFFICE + MANF - DISCBD + EMPACR + MPR

Ind/Off. Proj. HOTEL = RES + OFFICE + MANF - DISCBD + EMPACR + MPE

In these equations hotel development (1970) is a function of residential development (1970); office development (1970); manufacturing development (1970); distance to metropolitan CBD; employment density in the area of influence (base year); and in the case of the residential project, the size of the project and, in the case of the industrial project, the number of employees of the project. The reasons each of these variables were included are as follows:

a. RES, OFFICE, MANF

Residential development (1970), office development (1970) and manufacturing development (1970) were included as each of these generates overnight visitors and thereby increases the demand for hotel space. We can expect, therefore, that an increase in any of these uses will result in an increase in hotel development.

b. DISCBD

Hotel development tends to be most concentrated near population and employment centers. Therefore, as the distance to the CBD increases, we can expect a decrease in hotel development.

c. EMPACR

As indicated in both (1) and (2) above, hotel development is related to employment opportunities; therefore, there should be a relationship between the employment density in the area of influence in the base year and total hotel development in 1970.

d. MPR

In the residential project model the number of units of the residential project will result in an increased demand for hotel development and thereby an increase in the amount of hotel development.

e. MPE

In the industrial/office project the major project will result in an increased demand for hotel development from visitors who have business with the project. We can expect, therefore, that as the size of the project increases the amount of hotel development also increases.

The variables in the above equation relate hotel development to population and employment centers. Another factor influencing hotel development is proximity to major highways. Therefore, a potential variable in this equation is numbers of highway interchanges. In calibrating the model this variable will be considered in lieu of distance from the CBD.

7. Equation 7

Res. Proj. HOSPTL = RES - DISCBD + NONHSE + MPR

Ind/Off. Proj. HOSPTL = RES - DISCBD + NONHSE + MPE

In these equations hospital development is a function of the residential development (1970); distance to the CBD (base year); the nonhousehold density (base year); in the case of the residential project, the number of units of the major project and, in the case of the industrial/office project, the number of employees of the major project. The reasons each of these variables were included are as follows:

a. RES

Residential development was included as it generates a demand for hospital services. Thus, as residential development increases the amount of hospital development should also increase.

b. DISCBD

Hospital development services population and employment centers; therefore, as the distance to the CBD increases we can expect hospital development to decrease.

c. NONHSE

Nonhousehold population includes people in nursing homes and mental institutions. We can expect, therefore, a positive relationship between the nonhousehold density in 1960 and hospital development in 1970.

d. MPR

In the residential project model the number of units of the major project will increase the demand in the area for hospital development and thus, may result in an increase in such uses.

e. MPE

In the industrial office project the number of employees of the major project will increase the demand in the area for hospital development and thus, may result in an increase in such uses.

8. Equation 8

CULTUR = RES - DISCBD + MPR
CULTUR = RES - DISCBD

In these equations cultural facilities are a function of the residential development (1970); distance to the CBD (base year); in the case of the residential project, the number of units of the major project. The reasons each of these variables were included are as follows:

a. RES

Residential development generates a demand for cultural facilities. Thus, as residential development in an area increases, we can expect cultural facilities to also increase.

b. DISCBD

Land use studies have shown that cultural facilities tend to locate near population and employment centers. We can expect, therefore, that as the distance to the CBD increases, the amount of cultural facilities decreases.

c. MPR

As indicated in (1) above, residential development generates a demand for cultural facilities. We can expect, therefore, that the major residential project will likewise positively affect cultural facilities.

9. Equation 9

Res. Proj.	CHURCH = RES + MPR + MPKIDS
Ind/Off. Proj.	CHURCH = RES

In these equations, church facilities (1970) are a function of residential development (1970); in the case of the major residential project, the number of units of the project and the density of school age children. The reasons each of these variables were included are as follows:

a. RES

Residential development generates demand for church facilities. Thus, as residential development increases, church facilities will also increase.

b. MPR

As indicated in (1) above, residential development generates a demand for church facilities. We can expect, therefore, that the major residential project will likewise positively affect church facilities.

c. MPKIDS

The patronage of many churches is oriented towards families. We can expect, therefore, that an increase in the number of school age children of the major project may positively affect church facilities.

10. Equation 10

Res. Proj. EDUC = RES + MPR + MPKIDS + ENRACR
Ind/Off. EDUC = RES + ENRACR

In these equations public educational facilities (1970) are a function of residential development (1970); the density of public school enrollment (base year); and, in the case of the residential project, the number of units of the project and the density of school age children. The reasons each of these variables were included are as follows:

a. RES

Residential development generates a demand for educational facilities. Thus, it was included in our model as a prime factor influencing educational facilities.

b. ENRACR

Public school enrollment (base year) will have an affect on the amount of educational facilities (1970). Thus, this factor was included in our model.

c. MPR68 and MPKIDS

The residential major project will generate a demand for educational facilities. This demand will depend on the size of the project and the number of school age children per unit.

11. Equation 11

Res. Proj.	REC = RES + MPR + MINCC + INCMP
Ind/Off. Proj.	REC = RES + MINCC

In these equations,active outdoor recreation acres (1970) are a function of the residential development (1970); the income level in the area of influence (base year); and in the case of the residential project, the number of units of the major project (1970) and the income level of the major project (1970).

A potential substitute variable for MINCC is ENRACR and for INCMP is MPKIDS. The alternate theory is that the greater the number of children, the more active recreation acres the area of influence may have.

The reasons each of these variables were included in the model are as follows:

a. RES

Residential development generates a demand for active outdoor/active recreation acres. We can expect, therefore, that as the residential development in the area increases, the amount of active outdoor recreation area also increases.

b. MINCC

Recreation and land use studies by Metcalf & Eddy have shown that the income level in a community affects the amount (as well as the type) of active outdoor recreation acres. Therefore, we included income level in our model as a causal influence on recreation acres.

c. MPR + MINCC

The residential major project will increase the demand for active recreation acres in the area. This increase in demand will depend on the size of the project and on the income level of the project.

12. Equation 12

Res. Proj. $\mathrm{HWLM} = \mathrm{RES} + \mathrm{COMM} + \mathrm{OFFICE} + \mathrm{MANF} + \mathrm{HWYINT} - \mathrm{DISCBD} + \mathrm{EMPACR} + \mathrm{AUTO} + \mathrm{MPR}$

Ind. Proj. $\mathrm{HWLM} = \mathrm{RES} + \mathrm{COMM} + \mathrm{OFFICE} + \mathrm{MANF} + \mathrm{HWYINT} - \mathrm{DISCBD} + \mathrm{EMPACR} + \mathrm{AUTO} + \mathrm{MPE}$

In these equations highway facilities (highway lane miles) (1970) are a function of residential development (1970); commercial development (1970); office development (1970); manufacturing development (1970); the number of highway interchanges (base year +5); distance to the CBD; the density of employees/acre (base year); automobile drivers/acre in the country (base year); and in the case of the residential project, the number of units of the project (1970) and, in the case of the industrial/office project, the number of employees of the project (1970). The reasons each of these variables were included in the model were as follows:

a. RES, COMM, MANF, OFFICE

Each of these types of development may generate substantial vehicular traffic. We can expect, therefore, that additional highway facilities may be constructed in response to their needs.

b. HWYINT

Number of highway interchanges were included in the model as we felt that construction of a highway interchange would induce additional highway lane miles.

c. DISCBD

Highway facilities (highway lane miles) link outlying areas to population and employment centers. We can expect, therefore, an inverse relationship between highway facilities and distance to the CBD.

d. EMPACR

As indicated above, highway facilities serve employment centers. We can expect, therefore, that as the density of employees per acre increases, highway facilities also increase.

e. AUTO

This variable was included as an instrumental variable to break the loop between highway facilities and residential, commercial, manufacturing, and office development. The theory in our model is that the density of auto drivers in the county (1960) affects facilities in the area of influence in 1970.

f. MPE AND MPR

Each of these variables was included as the major project, and, in both cases, contributes to the amount of vehicular traffic generated in the area. This vehicular traffic needs highway facilities and thus, is a causal influence affecting such facilities.

IV. SAMPLE SELECTION

A. PURPOSE AND INITIAL CRITERIA

The purpose of the sample selection process was to identify for each type of major project a sample of case studies which could be used in the testing and calibration of the model. The criteria for the case studies were established jointly by the EPA contract officer and Walden Research. These criteria were:

- There would be twenty observations in each of the two samples (residential projects and industrial/office projects).
- No more than two case studies (observations) of each project type would be located in the same SMSA (Standard Metropolitan Statistical Area) and no more than three be located in the same Federal Region.
- The major project would be (a) a residential project, planned unit development, or new town with a minimum population of 4,500 or 1,800 parking spaces, or (b) an office or industrial park or research and development complex with a minimum enployment of 2,250 or 1,800 parking spaces.
- The project be constructed between 1954-1964 and attain at least 80 percent of the minimum size noted above within two years of initial operation.

The rationale behind the foregoing criteria was as follows:

- The location requirements were to assure representation throughout the United States.
- The size requirements were to assure that the project was "major".
- The timing requirements (i.e., construction dates and occupancy period) were to assure that project impacts had stabilized by 1970, the date in which land use in the area of influence was to be measured.

B. METHODOLOGY

To identify case studies which met the criteria, the following methodology was employed:

- Professional organizations which either may know of projects meeting the criteria or may be aware of sources for such projects were contacted. These organizations included the Urban Land Institute, American Society Planning Officials, and the National Association of Homebuilders. The organizations were helpful but did not have lists of potential case studies.
- Through the EPA project officer a questionnaire was sent to approximately 360 regional planning agencies requesting their assistance in identifying projects which met the criteria. This questionnaire is attached as Exhibit 4-1. The responses to the questionnaire identified approximately 75 potential projects. Subsequent investigation indicated that many of these projects did not meet the criteria and, thus, were eliminated from consideration. Nevertheless, the questionnaire proved to be the best source of potential case studies.

A telephone interview effort was undertaken. The objectives of this effort were:

- Contacting agencies who did not respond to the questionnaire for the purpose of identifying additional projects which met our criteria;
- Supplementing the information received to date concerning potential projects.

C. REVISED CRITERIA

After several weeks of attempting to identify qualified case studies, it became clear that the initial criteria established were too restrictive and that the list of qualified projects would be significantly extended, by relaxing these criteria. The particular criteria which many of the identified projects failed to satisfy were the criteria of 80 percent occupancy of the minimum size within two years and a construction period between 1954-1964. These criteria proved extremely difficult for residential projects where the majority of projects identified were developed and occupied over an extended period of time. For these projects, construction and occupancy were often both prior to and subsequent to 1965 and depended on market demand.

The changes to the initial criteria evolved from the problems encountered in finding acceptable case studies. Then changes were as follows:

- Phased development projects were permitted provided they met minimum size criteria. For residential projects the minimum size was reduced to 1000 units within 5 years.
- Ongoing projects were considered eligible provided substantial development occurred between 1960 and 1970.

This revised criteria permitted much greater flexibility in idenifying suitable case studies.

D. SELECTION PROCESS

Once a list of qualified case studies was prepared, the actual selection of the final sample took place. This selection process involved consideration of factors such as availability of information, particularly aerial photographs, and geographic location of the project. In addition, two case studies had to be replaced when problems became evident during the data collection task. The final list of case studies is shown for the industrial/office and residential sample on Table 4-1.

TABLE 4-1

CASE STUDIES

Industrial	Residential
Farmington Park, Farmington, CT	Joppatown, Hartford County MD
Western Electric, North Andover, MA	Montgomery Village, MD
Avco, Wilmington, MA	Kings Park, Fairfax County, VA
IBM, Kingston, NY	Vienna Woods, Fairfax County VA
Ft. Washington, Philadelphia, PA	Deltona, FL
Keystone Park, Scranton, PA	Miami Lakes, Miami, FL
Crestwood Park, Wright Twp., PA	Town'n Country, Tampa, FL
GE, Salem, VA	Montclair-Starmount, Charlotte, NC
Cummings Park, Huntsville, AL	Weathersfield, Schaumberg, IL
IBM, Lexington, KY	Oak Park, Blaine, MN
Collins Radio Park, Cedar Rapids, IA	Cottage Grove, MN
Ford, Woodhaven, MI	Clear Lake City, Harris County, TX
Western Electric, Columbus, OH	Meyerland, Houston, TX
White-Westinghouse, Columbus, OH	Westwood Heights, Omaha, NE
Little Rock Industrial Park, Little Rock, AR	Northglenn, CO
Chrysler, Fenton, MO	Sun City, Maricopa, AZ
Western Electric, Omaha, NE	Foster City, CA
Motorola, Phoenix, AZ	Huntington Harbour, Orange County, CA
Tektronix, Washington County, OR	Sun City, Perris, CA
Honeywell, Phoenix, AZ	Rancho Bernado, CA

EXHIBIT 4-1

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

Executive Director
Regional Planning Agency
Someplace, U.S.A.

Dear Sir:

The Land Use Planning Office is presently developing a methodology to predict induced land uses and resulting air quality impacts from major (1) housing facilities, planned unit developments, or new towns; and (2) office or industrial parks or research and development complexes. Twenty case studies are needed for each type of project to develop the model, which will be relatively easy to use and can assist practicing planners in environmental impact analyses, community facilities impact review, and air quality maintenance planning.

We would appreciate your assistance in helping us identify major projects in your region which might serve as case studies. Our basic criteria are:

1. The major project must be (a) a residential project planned unit development or new town with a minimum population of 4,500 or 1,800 parking spaces, or (b) an office or industrial park or research and development complex with a minimum employment of 2,250 or 1,800 parking spaces.
2. The project must be constructed between 1954-1964 and must attain at least 80 percent of the minimum size noted above within 2 years of initial operation.

If there are any projects within your region which meet or almost meet these criteria, would you please identify each project on the attached sheet and return it in the enclosed envelope by August 14.

Our consultants for this project are Walden Research and Metcalf & Eddy of Boston, Massachusetts. Representatives from these firms may be contacting you later when the final selection of cases is made.

Thank you very much for your assistance.

Very truly yours,

Thomas mc Curdy

Thomas McCurdy
Community Planner
Land Use Planning Office

EXHIBIT 4-1 (CONTINUED)

Name of Respondent: Organization:
Address: Telephone #:

Name of Project(s):
Type of Project (check one):
Residential ___ P.U.D. ___
Office Complex ___ Industrial Park ___ Research and Development ___
Location of Project (municipality):

Approximate Size of Project.
Resid., PUD, New Town - #DU's: Acres:
Office, Ind. Park, R&D Complex: Bldg Area (Φ): Acres:
Initial Year of Operation:
Approximate % Occupied within 2 Years:

Project Characteristics: Residential, PUD, New Town (Check Off):

Single Family Attached ___ Single Fam. Detached ___ Multifamily ___
Sales Housing ___ Rental Housing ___
High Income ___ Middle Income ___ Low Income ___
Predominately: Elderly ___ Family ___ Singles ___

Project Characteristics: Office, Ind. Park, and R&D (check off):
Finance Offices ___ Insurance Offices ___ Gov't Offices ___
Other Offices (specify):
R&D Developments (specify type):
Industrial Park Development (specify type):
Available Data Sources in Area for Planning Information.
3C Planning Agency:

Detailed Land Use Inventory of Project Area:

Please attach a map to any scale of the area encompassing the project(s) mentioned above. Feel free to make any pertinent comments concerning this project here. If you want to talk about this project, call Tom McCurdy at 919-688-8146 X291. Thanks for your assistance.

V. DATA COLLECTION

Following the specification of the model, a list of data items required for the model was prepared. This was supplemented with additional items potentially useful in the model calibration phase of this project. A list of the data items is shown on Table 5-1.

The data collection process consisted of two simultaneous phases. The first was an on site visit which primarily consisted of interviews with the local and regional planning agencies, the developer (if available), and the collection of locally available data. This data consisted primarily of building floor area of various categories of land uses in the area of influence, obtained by aerial photograph interpretation. The second phase consisted of the the collection of 1960 and 1970 Census of Population and Housing data for the area of influence. Field data collection forms for both phases are shown on Exhibit 5-1.

The data collection effort was preceeded by a test-training case. This case, the Wilmington, Massachusetts, AVCO Corporation Facility, was used to test the feasibility of collecting the requisite data, the clarity of the field data collection forms, and to train the personnel involved in the data collection effort.

The element of the data collection process that was most subject to potential measurement error was the aerial photograph interpretation. The classification scheme for the various land uses did not facilitate interpretation and some of the data collection personnel had little prior experience in aerial photograph interpretation. Three of the case studies had alternate sources of the same information which permitted an independent check on the process.

A. WHITE WESTINGHOUSE AND WESTERN ELECTRIC, COLUMBUS, OHIO

The Mid-Ohio Regional Planning Commission (MORPC) had a transportation planning data base which includes an inventory of building floor area

TABLE 5-1
LIST OF DATA COLLECTION ITEMS

Data Item	Definition	Aggregation	Time
MAJOR PROJECT:			
Total Housing Units		Major Project	1970
Housing Units		Major Project	T+0, T+2, T+8, 1960, 1970
Single family attached	Single family dwelling units that are attached such as row houses or townhouses		
Single family detached	Single family dwelling units which do not have common walls		
Multifamily low rise	Apartment units of 4 stories or less		
Multifamily high rise	Apartment units of more than 4 stories		
Economic level of residents		Major Project	T+2
Above average	Income of residents is approximately 15% or more above median for municipality		
Average	Income of residents is no more than 15% above or below median for municipality		
Below average	Income of residents is approximately 15% or more below median for municipality		
School children per unit	Average school age children per unit	Major Project	T+2
Number of acres	Total developed acres and undeveloped acres under developer control	Major Project	1970
Type of industry	% Manu, % Office, % R & D	Major Project	1970
Total Employment	Total employees working at the project site	Major Project	T+0, T+2, 1960, 1970
Percent Professional	Managerial and professional	Major Project	1970
ENDOGENOUS VARIABLES:	Square feet of the following; land use codes as used by the Public Administration Service in its 1962 Land Use Classification Manual	Area of Influence*	1970
Office	LUC 60 Finance, Insurance, Real Estate 62 Business Services 67 Medical, Health, Legal Services 68 Other Professional Services		
<50 thousand square feet (<4.65x10^3 sq. meters)			
50-100 thousand square feet (4.65x10^3 to 9.29x10^3 sq. meters)			

TABLE 5-1 (Continued)

LIST OF DATA COLLECTION ITEMS

Data Items	Definition	Aggregation	Time
>100 thousand square feet ($>9.29\times10^3$) sq. meters			
Commercial <25 thousand square feet ($<2.32\times10^3$ sq. meters) 25-50 thousand square feet (2.32×10^3 to 4.65×10^3 sq. meters) 50-100 thousand square feet (4.65×10^3 to 9.29×10^3 sq. meters) >100 thousand square feet ($>9.29\times10^3$ sq. meters)	LUC 52-59 Retail Trade 61 Personal Services 63 Automobile Service 64 Miscellaneous Repair Service 65 Indoor Amusement Services	Area of Influence	1970
Manufacturing	LUC 2 Nondurable Goods Manufacturing 3 Durable Goods Manufacturing	Area of Influence	1970
Wholesale/Warehousing	LUC 50 Wholesale 46 Warehousing	Area of Influence	1970
Hotels/Motels <25 thousand square feet ($<2.32\times10^3$ sq. meters) 25-50 thousand square feet (2.32×10^3 to 4.65×10^3 sq. meters) 50-100 thousand square feet (4.65×10^3 to 9.29×10^3 sq. meters) >100 thousand square feet ($>9.29\times10^3$ sq. meters)	LUC 07 Hotels, Motels, Tourist Homes	Area of Influence	1970

TABLE 5-1 (Continued)

LIST OF DATA COLLECTION ITEMS

Data Item	Definition	Aggregation	Time
Churches	LUC 764 Churches 765 Other Religious Services	Area of Influence	1970
Cultural	LUC 76 Museums, Libraries, Art Galleries, except Churches (764), Arboreta (762), Cemeteries (767)	Area of Influence	1970
Hospitals <50 thousand square feet ($<4.65\text{x}10^3$ sq. meters) 50-100 thousand square feet ($4.65\text{x}10^3$ to $9.29\text{x}10^3$ sq. meters) >100 thousand square feet ($>9.29\text{x}10^3$ sq. meters)	LUC 77 Hospitals, Sanatoria, Convalescent Homes and Rest Homes	Area of Influence	1970
Public Education <50 thousand square feet ($<4.65\text{x}10^3$ sq. meters) 50-100 thousand square feet ($4.65\text{x}10^3$ to $9.29\text{x}10^3$ sq. meters) >100 thousand square feet ($>9.29\text{x}10^3$ sq. meters)	LUC 74 Public Schools 75 Private Schools	Area of Influence	1970
Active Outdoor Recreation	LUC 80 Outdoor Public Recreation 81 Outdoor Water Based Recreation	Area of Influence	1970

TABLE 5-1 (Continued)
LIST OF DATA COLLECTION ITEMS

Data Item	Definition	Aggregation	Time
EXOGENOUS VARIABLES			
Vacant Land	Vacant developable acres: vacant land not lying within wetland area as shown on U.S.G.S. Map, nor having slope greater than 15%, nor part of the permanent open space system.	Area of Influence	T+0
DU/Acre	Total dwelling units for the municipality or tract as presented in Table H-1 of the 1960 Census Tract data or Tables 12, 18, 22, 25 or 27 of the 1960 General Housing Characteristics divided by total acres of the respective area.	Municipality or Census Tract	1960
Percent vacant housing units	Available vacant units divided by all housing units as provided on Table H-1 in the 1960 Census Tract data or total housing units less occupied units as presented in Tables 12, 18, 24, or 26 of the 1960 General Housing Characteristics.	Municipality or Census Tract	1960
Highway Interchanges	Limited access highway interchanges both full and partial as depicted on U.S.G.S. Map or air photos.	Area of Influence	T+5
Median Income of Families and Individuals	As listed in Table P-1 of the 1960 Census Tract Data or Tables 76 or 81 in the 1960 General Social and Economic Characteristics.	Municipality or Census Tract	1960
Office Vacancy	Percent of all office buildings vacant as presented in the survey by the Building and Office Managers Association.	Metropolitan Area	T+0
Financial Employment per acre	Employment for Standard Industrial classification for Finance, Insurance, and Real Estate for Municipality divided by acreage of the municipality.	Municipality	T+0
Distance to CBD	Highway distance from center of major project to center of the major central business district as defined by the U.S. Bureau of the Census.	Metropolitan Area	T+0
Energy Cost	Dollars per kilowatt hour for commercial users as presented by data from the Federal Power Commission.	Metropolitan Area	T+0
Miles of Railroad	Linear miles of active freight and passenger railroad lines	Area of Influence	T+0

TABLE 5-1 (Continued)

LIST OF DATA COLLECTION ITEMS

Data Item	Definition	Aggregation	Time
Warehouse and Wholesale Employment per acre	Employment with SIC 422, 50 and 51 as reported by the State Division of Employment Security divided by total acres of the municipality or district.	Municipality	T+0
Total Employment per acre	Total employment as reported by the State Division of Employment Security divided by the total acres of the municipality or district.	Municipality	T+0
Manufacturing Employment	SIC 20-39.	Municipality	T+0
Non-Household Population per acre	Total population less population in households as presented in Table P-1 of the 1960 Census Tract data or Tables 15, 20, 24, 26 or 27 of the 1960 Census Urban places, towns and Minor Civil Divisions, depending on the size of the municipality divided by total acres of the respective delineation.	Municipality or Census Tract	1960
Enrollment per acre	Total public school enrollment divided by total acres of the respective municipality or school district.	Municipality	T+0
Regional Population	Total population for the counties as presented in the 1960 and 1970 U.S. Census General Population Characteristics.	Counties in Region	1960, 1970
Regional Employment	Total employment for the counties as reported by the State Division of Employment Security (adjusted to represent a similar number of covered professions).	Counties in Region	1960, 1970
Regional Median Income	As reported by Census.	Counties in Region	1960
Zoning Classification % Residential % Single Family Acres % Multifamily % Office % Commercial % Industrial % Other	As indicated in local zoning ordinance	Area of Influence	T+5

TABLE 5-1 (Continued)

LIST OF DATA COLLECTION ITEMS

Data Item	Definition	Aggregation	Time
Professional Labor Force	Percent of total employed population, male and female, with occupations as professional, technical and kindred workers plus managers, officials, and proprietors as presented in Table P-3 of the 1960 Census Tract Data or Tables 74 or 78 in the 1960 General Social and Economic Characteristics.	Municipality or Census Tract	1960
School-aged children per acre	Total school-aged children for municipality of census tract divided by acres of the respective district.	Municipality or Census District	T+0
Population in Area of Influence	As defined in Table HC 3 Block Statistics 1970 U.S. Census.	Area of Influence	1970
Total year round units	As defined in Table HC 3 Block Statistics 1970 U.S. Census.	Area of Influence	1970
One Unit Structures	As defined in Table HC 3 Block Statistics 1970 U.S. Census.	Area of Influence	1970
Highway Lane Miles	Lanes times miles of highways carrying over 10,000 average daily traffic.	Area of Influence	1970
Percent served by public sewer	Percentage of land which is connected to the public sewer or has public sewerage available.	Municipality	T+5
Commuter rail stops	Number of commuter railroad stops.	Area of Influence	T+0
Mass Transit stops	Number of fixed rail mass transit stops.	Area of Influence	T+0
Distance to airport	Highway distance to major airport (with FAA regulated tower).	Metropolitan Area	T+0
Presence of Agricultural uses University Private school Airports 5 mile square water body (12.95 sq. kilometers)	Existence of item within 3.23 miles (5.2 kilometers) of center of major project.	Metropolitan Area	T+0

Note: T = Base Year

EXHIBIT 5-1
PART A

ON SITE VISIT
DATA COLLECTION FORMS

EXHIBIT 5-1

Major Project Name____

Location (City & State)____________________________________

Location (UTM coordinates)___☐☐☐.☐___☐☐☐☐.☐______

Horizontal KM Vertical KM

USGS quadrangle sheets for area of influence (7.5 minute series).

1.__________________________________

2.__________________________________

3.__________________________________

4.__________________________________

Definition of area of influence

☐ Circle

☐ Other (describe)

Are there Non-Tracted Areas in the Area of Influence?

☐ No

☐ Yes (Indicate on Map)

EXHIBIT 5-1 (CONTINUED)

I. MAJOR PROJECT INFORMATION

Name of Project ______________________
City of Major Project ______________________
Exact Location ______________________
Contact Name ______________________
Contact Telephone Number ______________________
Developed Acres _ _ _ _ _ _ _ _ _ _ _ _ _ _ 1.
Undeveloped Acres Under Control of Developer, contiguous, 1970 _ 2.
Year of Initial Occupancy (Base Year) _ _ _ _ _ _ _ 3.

RESIDENTIAL PROJECT:

	Initial Year	Initial Year +2	1960	1968	1970
Number of D.U.	4.	5.	6.	7.	8.

	SFD	SFA	Low Rise Multifamily	High Rise Multifamily
Average sq.ft.,1970	9.	10.	11.	12.
1970 # of D.U.	13.	14.	15.	16.

ECONOMIC LEVEL OF RESIDENTS, T+2 - Median Income is

Above Surrounding Community (15%) _ _ _ _ _ _ _ _ 17.
About the Same _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ 18.
Below that of Surrounding Community (15%) _ _ _ _ 19.
Average Children Per Unit, T+2 _ _ _ _ _ _ _ _ _ _ 20.

INDUSTRIAL/OFFICE/RESEARCH PROJECT:

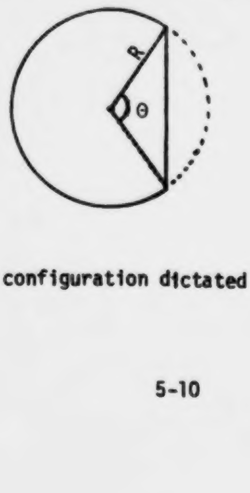
		PCT Manuf.	PCT R & D	PCT Office	Other
Type of Industry:	Percentage	21. %	22. %	23. %	24. %
		Initial Year	Initial Year +2	1968	1970
Employment		25.	26.	27.	28.

II. AREA OF INFLUENCE - 10,000 Acres, 15.625 square miles.

☐ 1. If no Coastline, Barriers, etc.

Circle of 2.23 miles radius   Circle of 2.23 miles radius

☐ 2. Otherwise, use circle without appropriate segment



15.625 = Area of circle - Area of Segment

$$15.625 = \pi R^2 - \frac{1}{2} R^2 (\theta - \sin\theta)$$

∴ necessary R for 10,000 acres is

$$\frac{1}{R} = \sqrt{\frac{\pi - \frac{1}{2}(\theta - \sin\theta)}{15.625}}$$

N.B. (θ in radians)

☐ 3. Or use other appropriate configuration dictated by situation. (Describe).

EXHIBIT 5-1 (CONTINUED)

Project name________________________________

III. NON-RESIDENTIAL BUILDING FLOOR AREA IN AREA OF INFLUENCE

1970 Floor Area in Thousands of Square Feet

	<25K	25-50K	50-100K	>100K	SOURCE
					1970 Air Photos Checked with 1970 Land Use Map and 1970 U.S.G.S. Map.
Commercial	29.	30.	31.	32.	SIC: 52-59,72,75,76,78,79 LUC: 52-59,61,63,64,65
Hotels & Motels, Clubs	33.	34.	35.	36.	SIC: 70 Hotel,Motel Redbook - 1970
		<50K	50-100K	>100K	
Office		37.	38.	39.	SIC: 60-67,73,80,81,89,86 LUC: 60,68,62,67,78
Hospital		40.	41.	42.	SIC: 80 (Interview with & AHA LUC: 77 local planner) Guide
Public Schools		43.	44.	45.	SIC: 82 LUC: 74,75
Manufacturing		46.			SIC: 20-39 LUC: 2,3
Wholesale & Warehousing		47.			SIC: 50,51,422 LUC: 50,46
Churches		48.			SIC: 866 LUC: 764
Cultural		49.			SIC: 84 (Interview with local LUC: 76 (except 764) planner).

LAND USE IN AREA OF INFLUENCE

**Active Outdoor Recreation, terminal year 50. , acres.

**Includes, playgrounds, playfields, tot lots, golf courses, riding stables, developed beaches, tennis complexes and other active recreation such as ski slopes.

Does not include, passive open space portions of large reservations, arboretums, other areas devoted to conservation, such as, wildlife refuge, national forests.

The inventory of active outdoor recreation is an optional item. Do only if you have the time or is easily available from secondary information.

INDICATE PRESENCE OF OTHER LAND USES IN 1970

	Universities		Private Schools		Airports		Coastline		Inland Water 5 acres +		Agriculture	
Within area of Influence	51.	52.	53.	54.	55.	56.	57.	58.	59.	60.	61.	62.
Within 1 Mi. of Boundary of Area of Influence	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No

Describe any other significant land uses or factors potentially affecting development (continue on reverse side). (i.e., military bases, correctional institutions, etc.).

EXHIBIT 5-1 (CONTINUED)

Project name ____________________

IV. GEOGRAPHIC DATA:

1970 Highway Lane Miles with Road A.D.T. > 10,000 (from worksheet #2) 74.

Area of Influence:

Highway Interchanges t + 5 --- 75.
Commuter Rail Stops t 0 --- 76.
Mass Transit Stops t 0 --- 77.
Railroad Line Miles t 0 --- 78.

Major Project:

Distance to Region CBD t 0 ------------------------ Highway miles ------ 79.
Distance to F.A.A. Airport t 0 ------------------- Highway miles ------ 80.

V. MUNICIPAL DATA:

For Each Principal Municipality (if not base year, then you must adjust for presence of Major Project). Obtain vacant land, zoning, and sewer information for area of influence if possible.

Name of Municipality or Area				
Total Area in Acres	81.			
Vac. Developable Acres, t+0	82.			
Vac. Undevelopable Acres, t+0	83.			
Percent sewered t+5	84. %	%	%	%
Percent of Area Zoned in following categories, t+5				
Single family residential	85. %	%	%	%
Multifamily residential	86. %	%	%	%
Commercial	87. %	%	%	%
Office	88. %	%	%	%
Industrial	89. %	%	%	%
Other	90. %	%	%	%
Total Employment t + 0	91.			
Manufacturing Emp. (SIC 20-39)	92.			
Office Emp. (60-67,73,80,81,86,89)	93.			
Wholesale & Warehouse Emp (50,51)	94.			

VI. REGIONAL DATA:

For Each County Economically Significant to Area of Inf.

County				Census
Total Employment: 1960	95.			
1970	96.			
Area	97.			

SMSA/Metropolitan Area

Base Year: Cost of Energy, $/KWH t+0 98.
Office Vacancy Rate, percent 99. %
Per Capita Effective Buying Income 100.

EXHIBIT 5-1 (CONTINUED)

Project name________________________________

WORK SHEET HWLM with ADT >10,000

Link Name	Approximate ADT	Link Length	Link Lanes	Lane Miles

EXHIBIT 5-1 (CONTINUED)

COMMERCIAL				MANUFAC-TURING	WAREHOUSING WHOLESALING	OFFICE	10 HOTELS & MOTELS CLUB	11 HOSPITALS	12 PUBLIC SCHOOLS	13 CHURCHES	14 MISCEL-LANEOUS
<25K	25-50K	50-100K	>100K				<25K	<50K	<50K		
							25-50K				
								HOSPITAL 50-100K	SCHOOLS 50-100K		
							50-100K			CULTURAL	
							> 100K	HOSPITALS >100K	SCHOOLS >100K		

PROJECT NAME____________________

EXHIBIT 5-1

PART B

CENSUS OF POPULATION AND HOUSING DATA COLLECTION FORMS

EXHIBIT 5-1 (CONTINUED)

Project name________________________________

1970 CENSUS HC(3) BLOCK STATISTICS, TABLE AVAILABLE FOR ALL URBANIZED AREAS.
IF NOT COVERED, HOUSING COUNTS TO BE TAKEN FROM AERIAL OR OTHER MEANS
FOR ALL BLOCKS WHOLELY OR PARTIALLY IN AREA OF INFLUENCE

CENSUS TRACT	BLOCK NUMBER	% OF BLOCK IN AREA	TOTAL POPULATION	TOTAL YEAR ROUND UNITS	ONE UNIT STRUCTURE
		TOTAL:	101.	102.	103.

EXHIBIT 5-1 (CONTINUED)

Project name______________________________

1970 CENSUS PHC(1) CENSUS TRACTS, TABLES H-1 & H-2
AVAILABLE FOR ALL SMSA S AND MORE. IF NOT COVERED, INFORMATION
IS COLLECTED BY PRINCIPAL MUNICIPALITIES FOR ALL CENSUS TRACTS MORE THAN 25%
IN AREA OF INFLUENCE AND OTHER CENSUS TRACT JUDGED TO BE RELEVANT

							Total
Census Tract No.							
% of Census Tract in Area							
Median Rooms							104.
Year Round Units							105.
Units	1						106.
in	2						107.
Structure	3-4						108.
	5-49						109.
	50+						110.
Occupied Units							111.

If area of influence wholely or partially in untracted area without block statistics, enter estimate of 1970 housing units for untracted/unblocked area. 112. Units, year round.

EXHIBIT 5-1 (CONTINUED)

Project name________________________

FOR ALL PRINCIPAL MUNICIPALITIES [All Municipalities Within One-Half Mile of Project]

1970 CENSUS DETAILED HOUSING CHARACTERISTICS

GENERAL HOUSING CHARACTERISTICS

ITEM	TABLE NO. BY SIZE OF MUNICIPALITY				DATA			TOTAL
	50,000 Plus	10,000-50,000	10,000-2,500	1,000-2,500				
Name								
% of City in Circle					%	%	%	%
Year Round Units	D45	D55	D58	G27				114.
Elevator: Total Units					In Towns Less Than 10,00, Estimate No. of High Rise			
4 Floors +	D45	D55	-	-				115.
1-3 Floors	D45	D55	-	-				116.
Units in Structure:					In Towns 10-2 1/2 Thousand, Estimate Detached/Attached Singles. In Towns Less Than 2,500, One Units are Detached, Remainder 3-4 Units.			
1 Detached	D43	D53	-	G27				117.
1 Attached	D43	D53	-	-				118.
2	D43	D53	D58	-				119.
3-4	D43	D53	D58	G27				120.
5 +	D43	D53	D58	-				121.
Mobile Home	D43	D53	D58	-				122.

FILL OUT FOLLOWING SECTION FOR UNTRACTED - UNBLOCKED PRINCIPAL MUNICIPALITIES

ITEM					DATA			TOTAL
Occupied Units					Towns Less Than 10,000, Add Renter & Owner Occupied			
All	G8	G18	G23	G27				123.
Median Rooms/Unit					Towns Less Than 2,500 Take Weighted Average of Renter & Owner			
All Year Round	G9	G19	G23	G27				124.

FILL OUT FOLLOWING FOR COUNTIES IN URBANIZED AREA OF PROJECT

ITEM	TABLE NO.	DATA			TOTAL
County Name					
Total Population	G29				125.
Total Housing Units	G29				126.

EXHIBIT 5-1 (CONTINUED)

Project name________________

1960 CENSUS OF POPULATION & HOUSING

Use the combination of Census Tracts and minor civil divisions that best reflect the project area. If the same, use Census Tracts as the data is presented in a more usable format.

Note combination to be used:

Total Acreage of Census Tracts 113.

FOR CENSUS TRACTS

ITEM	TABLE	DATA				TOTAL
Tract No.						
Total Population	P-1					127.
Pop. in Household	P-1					128.
Median Income Family & Indiv.	P-1					129.
Pop. under 14	P-2					130.
Total Employment	P-3					131.
Professional	P-3					132.
Managers	P-3					133.
All Housing Units	H-1					134.
Owner Occupation	H-1					135.
Renter Occupation	H-1					136.
Available Vacant	H-1					137.
Median Rooms/Unit	H-1					138.
Median Value, Owner	H-2					139.

EXHIBIT 5-1 (CONTINUED)

Project name____________________________

FOR URBAN PLACES, TOWNS, & MINOR CIVIL DIVISIONS

It is important that all items in each block are from a common base. Minor Civil Division and Urban Place are not the same.

BLOCK 1 - POPULATION

ITEM	TABLE #	NAME OF TOWN			TOTAL
Total Population	25				140.
Number of Households	25				141.
Population in Households	25				142.
Males under 5	26				143.
Males 5-14	26				144.
Females under 5	26				145.
Females 5-14	26				146.
Median Income Family & Indiv.	76 or 81				147.
Employment					148.
Professional Employment	74, 78, or 81				149.
Managerial Employment					150.

BLOCK 2 - HOUSING

ITEM					
Total Housing Units	12 18 22 25 27				151.
Population in Housing Units	15 20 24 26 27				152.
Occupied Housing Units					
Total	12 18 24 26 --				153.
Owner	-- -- -- -- 27				154.
Renter	-- -- -- -- 27				156.
Mean-Median No. of Rooms					
Total	13 19 23 25 --				156.
Owner	-- -- -- -- 27				157.
Renter	-- -- -- -- 27				158.
Median Value-Owner Occupation	17 21 24 26 27				159.

Fill out for all Counties Economically Significant to the Area of Influence

Total Population	P-13				160.
Total Housing Units	H-28				161.
Total Workers	G-82				162.
Auto Drivers	G-82				163.
Median Income					164.

by zone. As such, it provides an estimate of the accuracy of the methodology that was employed to estimate floor area from aerial photographs.

The MORPC transportation zones for both case study areas are shown in Figure 5-1 and 5-2. The floor area of all the zones contained in the two case study areas was summed and compared to the figures obtained by Walden from the aerial photographs. The floor area in zones along the boundary of the circle was apportioned to the total by area. Table 5-2 shows the comparison for, respectively, the Western Electric and White-Westinghouse case studies. The total floor area estimated from the aerial photographs is remarkably close to the MORPC inventory; the difference is within two percent for Western Electric and 1.5 percent for the White-Westinghouse case. Due to the different classification systems used by MORPC and Walden, a comparison of the subtotals is more difficult.

From our knowledge of the study areas and discussion with personnel, we made the following observations:

- The office floor area was systematically underestimated. This is most likely due in part to misclassifying small offices as commercial floor space (e.g. bank branches in shopping centers, real estate offices, etc.). One other possibility is that auxiliary office space is being misclassified in industrial facilities as industrial floor area; however, we believe that the "missing office area is classified as commercial floor area. We anticipate that will cause no difficulty with the final product viz., predicted emissions.
- The amount of manufucturing floor area is overestimated by about ten percent. This could be due to misclassifying some warehousing as manufacturing; however, MORPC personnel consider that the distinction between non-manufacturing industrial floor area and manufacturing industrial floor area may not have been consistently accurate in their inventory.
- As the MORPC groups all retail and wholesale floor area and we grouped warehousing with wholesale floor area, it was not possible to accurately evaluate the comparison for these subtotals. Discussion with MORPC personnel familiar with the case study areas indicate that our figures are reasonable.

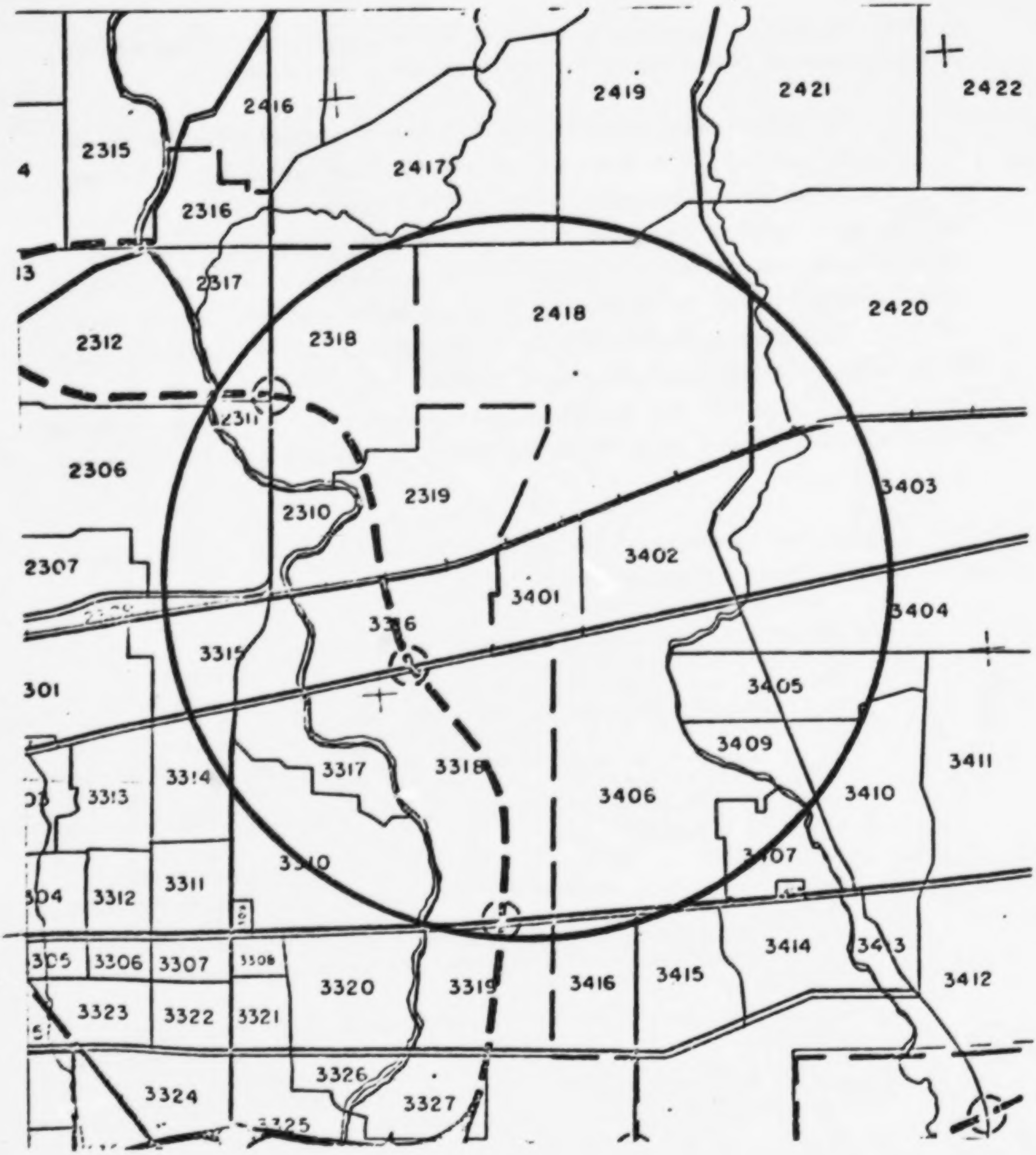

2315
2416
2419
2421
2422
2417
2316
2317
2418
2420
2312
2318
2311
2306
2319
3403
2310
2307
3402
3401
3404
3316
3315
301
3405
3409
3411
3317
3318
3314
3313
3406
3410
3407
3310
3311
3312
3414
3305
3306
3307
3308
3320
3319
3416
3415
3413
3412
3323
3322
3321
3326
3324
3327
3325

Figure 5-1 Western Electric Case Study Area and MORPC Transportation Zones

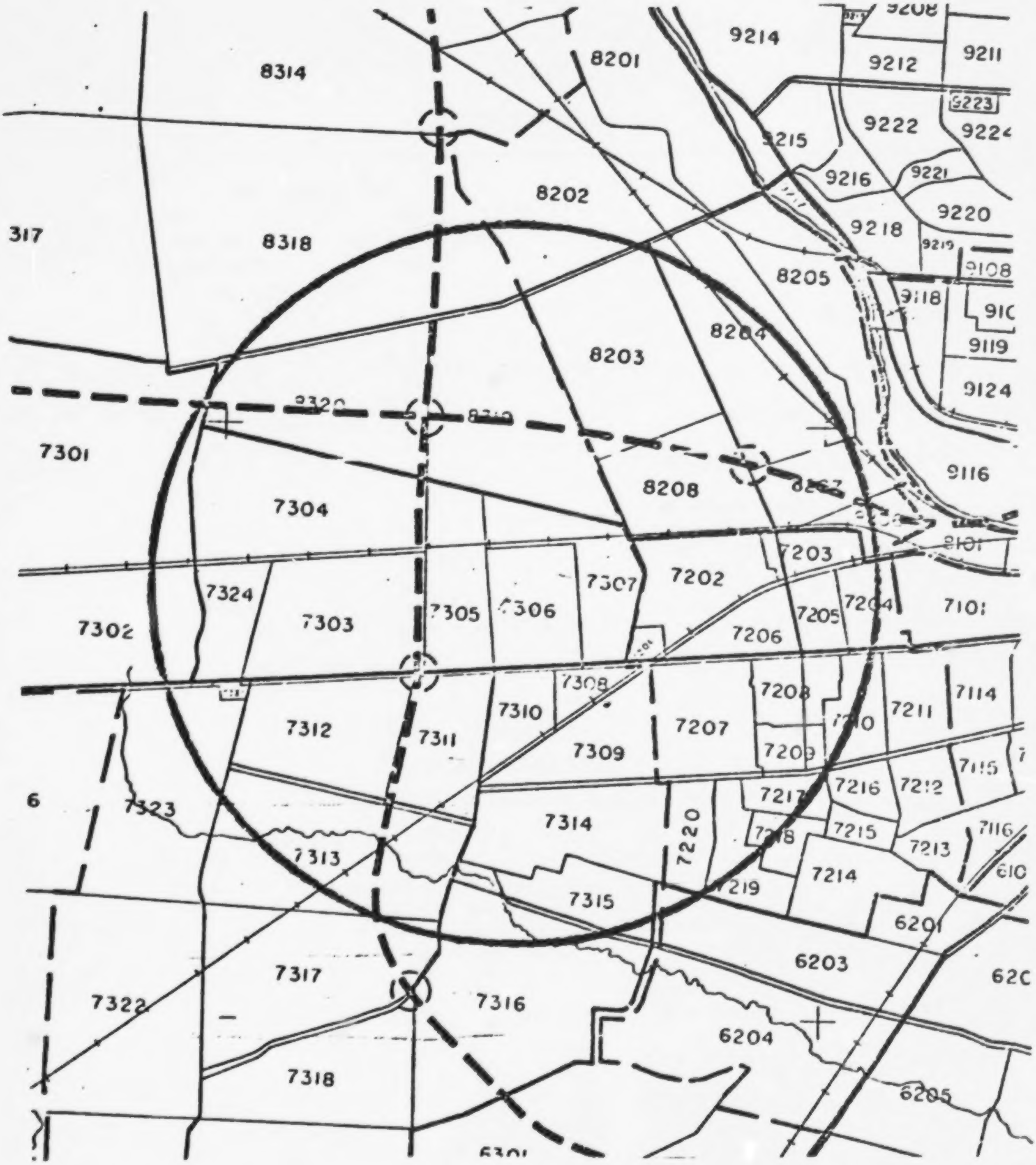

9208
9214
9211
8201
9212
8314
9223
9222
9224
9215
9216
9221
8202
9220
317
8318
9218
9219
9108
8205
9118
8204
8203
9119
9124
8320
7301
9116
7304
8208
8101
7203
7324
7307
7202
7302
7303
7305
7306
7205
7204
7101
7206
7308
7208
7114
7310
7211
7312
7311
7207
7309
7209
7115
6
7323
7216
7212
7217
7314
7220
7218
7215
7116
7213
7313
7214
7219
7315
6201
6203
7317
7322
7316
6204
7318
6205
6301

Figure 5-2. White-Westinghouse Case Study Area and MORPC Transportation Zones

TABLE 5-2

WESTERN ELECTRIC

	10^3 square feet MORPC Transportation Zones	10^3 square feet WALDEN by Aerial Photographs
Office	76	48
Retail and Wholesale Trade and Services	597	
Retail Trade and Services		568
Warehousing and Wholesale		5
Warehousing, Construction	252	
Manufacturing	2,372.6	2,601
TOTAL	3,279.6	3,222

TABLE 5-2

WHITE-WESTINGHOUSE

	10^3 square feet MORPC Transportation Zones	10^3 square feet WALDEN by Aerial Photographs
Office	362	238
Retail and Wholesale Trade and Services	5,360	
Retail Trade and Services		1,874
Warehouse and Wholesale Trade		6,010
Warehousing, Construction	2,617	
Manufacturing	4,158	4,572
TOTAL	12,497	12,694

B. MONTCLAIRE - STARMONT, CHARLOTTE, NORTH CAROLINA

A more limited check on the aerial photograph interpretation process was available on this case study. A land use data bank (Management Planning Analysis Module, MPAM) was available on a Census Tract disaggregation. A comparison of the aerial photogrpah interpretation and the MPAM data is shown on Table 5-3. The total floor area estimated from the aerial photographs is 78 percent of that reported in the MPAM data bank. There are also apparently some severe misclassification errors.

In summary, we are unable to adequately explain the discrepancy in the Charlotte, North Carolina aerial photograph interpretation. The three aerial photograph interpretations referenced in this section, were each done by different individuals. Of the three, the individual doing the interpretation in Charlotte was the most experienced. We have identified two possible reasons for the conflicting results between the three sites:

1. Photograph Scale

The Columbus, Ohio aerial photographs were 1" = 400' scale; the Charlotte, North Carolina photographs were 1" = 800' scale.

2. Data Aggregation

The MORPC data was on a smaller aerial unit than the Charlotte data. The MORPC data had about twenty zones in each circle (see Figures 5-1 and 5-2); the Charlotte data had eight. While the Charlotte MPAM data may be accurate, the aggregation and apportionment necessary to compare it with the totals from the aerial photograph interpretation may have caused the error.

The second reason would have no impact on the accuracy of this project. If the error was caused in part by the smaller scale of the Charlotte photographs, certain case studies may be more accurate than others. The distribution of the aerial photograph scales of all case studies is shown on Table 5-4.

TABLE 5-3

COMPARISON OF BUILDING AREA DATA (10^3 Sq. Ft.)

Item	Estimate from 1971/1972 Aerials	MPAM 1973 Data
Offices	1,489.4	2,532
Commercial	2,606.5	2,946
Industrial	2,835.1	1,654
Remainder	925.7[a]	1,262[b]
Total	6,516.7	8,394

[a] Includes: Hotels, Motels, Clubs; Hospitals; Public Schools; Wholesale and Warehousing; Churches; and Cultural.

[b] Includes: Institutional and other.

TABLE 5-4

CUMULATIVE PERCENTAGE OF CASE STUDY DATA COLLECTION OBTAINED FROM PHOTOGRAPHS OF CERTAIN SCALE OR LARGER

1" = 100	5%
1" = 200	12.5%
1" = 300	17.5%
1" = 400	35%
1" = 500	45%
1" = 600	57.5%
1" = 700	62.5%
1" = 800	75%
1" = 900	77.5%
1" = 1000	90%
1" = 1200	97.5%
1" = 2000	100%

VI. CAUSAL ANALYSIS

A. GENERAL APPROACH

The approach to path analysis in the current study involved the use of two basic statistical techniques: two stage least squares and ordinary least squares multiple regression. The first technique was used to solve for path coefficients in the system of five equations connected by feedback loops (viz, the dependent variables RES, COMM, OFFICE, MANF, and HWLM). For a given dependent variable, the first stage of the two stage process involved estimating the values of the other four endogenous variables through linear combinations of so-called instrumental variables which are chosen for their causal relationships with the endogenous variables. Take, for example, the Residential model equation where RES is the dependent variable. OFFICE, MANF, and HWLM are endogenous, and MPR70, DUACRE, and VACACR are exogenous:

$$\begin{aligned} \text{RES} = {} & b_1\ \text{OFFICE} + b_2\ \text{MANF} + b_3\ \text{HWLM} + b_4\ \text{MPR70} + \\ & b_5\ \text{DUACRE} + b_6\ \text{VACACR} + b_7 + \varepsilon \end{aligned}$$

If ordinary least squares was applied directly to this equation, the estimates of b_1, b_2, . . . b_7 would be inconsistent because three of the model variables (OFFICE, MANF, and HWLM) are endogenous variables inked to RES through feedback loops and hence correlated with the error term ε [35]. However, we can eliminate the dependence of the endogenous variables on ε by substituting modified regression estimates of OFFICE, MANF, and HWLM that are linear combinations of the exogenous (i.e., independent of ε) instrumental variables. First stage regressions are therefore performed with OFFICE, MANF, and HWLM as the dependent variables and the instrumental variables as the independent variables to obtain values for the modified regressors, $\overline{\text{OFFICE}}$, $\overline{\text{MANF}}$ and $\overline{\text{HWLM}}$. The second stage then entails ordinary least squares of the following equation:

$$\begin{aligned} \text{RES} = {} & b_1\ \overline{\text{OFFICE}} + b_2\ \overline{\text{MANF}} + b_3\ \overline{\text{HWLM}} + b_4\ \text{MPR70} + \\ & b_5\ \text{DUACRE} + b_6\ \text{VACACR} + b_7 + \varepsilon \end{aligned}$$

Because of the interconnections among the five equations, the set of instrumental variables used in any one, two stage regression had to consist of all of the instrumental variables in the simultaneous block. A complete discussion of statistical techniques is given in Section II, and the problem of identification that can occur in two stage least squares analysis is discussed in Section III.D.

An examination of the model equations (see Section VI.E.2) showed there to be 20 instrumental variables. For the Residential project, for example, these were:

MPRT2
MPR68
MPR70
DUACRE
VACACR
VACHSG
HWYINT
MINCC
INCMP
OFFVAC
OFFACR
DISCBD
MINCR
ENERGY
RRMI
EMPACR
MANACR
DELEMP
DELPOP
AUTO

The number of instrumental variables in the original model was a definite problem, since meaningful results cannot be obtained in the first stage regression if there are as many independent (instrumental) variables as there are data samples in the current study (20). Therefore, a modified stepwise regression analysis was performed to choose those instrumental variables which were most appropriate, in order to allow sufficient degrees of freedom for the error term in the first stage regression and ensure stable coefficients for the first stage estimates. This analysis and the trimmed equations are discussed in Section VI.C.4.

To solve for path coefficients in the other model equations that were not interconnected, ordinary least squares regression techniques were used.

The multiple regression techniques used in path analysis, when applied directly to the data to be analyzed, will yield unstandardized path regression coefficients (b). However, if applied to data that has been standardized (i.e., the data has a mean of 0 and a standard deviation of 1), the analysis will yield standardized path coefficients (β) or beta weights. There are advantages to both the standardized and the unstandardized approaches. Unstandardized path regression coefficients indicate the amount of change in a dependent variable given a one unit change in a particular independent variable, all other effects held constant. This approach allows one to deduce quantitatively the resultant effects of a given change or action on the system. On the other hand, standardized path coefficients provide a sensible way of comparing the relative importance of various independent variables independent of different units and scales. In the current study, both b and β values were computed for each regression performed. The path coefficients (β) were used extensively in judging the significance of model paths for theory trimming. Path coefficients do not, however, provide a valid basis for comparisons of systems operating on different data populations since the variance of the individual variables is included in the path coefficients and variance can change drastically from sample to sample [26]. Therefore, path regression coefficients (b) were utilized in the analysis of model stability. Both b and β values are presented in this report for the final model equations.

Two different computer software packages were employed for the statistical analysis in this study. The first of these was the Time Series Processor (TSP) developed at MIT and Harvard University in the late 1960's [39]. TSP is a programming language oriented towards the statistical analysis of time series, with specific applications to econometric research. TSP was used for all two stage least squares analysis. Although it also provides ordinary least squares techniques, these analyses were not performed on TSP due to several software installation problems encountered during the project. Therefore, the Statistical Package for the Social Sciences (SPSS) was used instead [40]. SPSS is a comprehensive and widely used system of programs for the statistical analysis of social science data. Although SPSS automatically provides both b and β values for a multiple regression, TSP does not. Thus, β values for the two stage least squares were calculated using the following identity:

$\beta_i = b_i s_i / s_y$

where β_i = path coefficient for dependent variable x_i

b_i = path regression coefficient for dependent variable x_i

s_i = standard deviation of independent variable x_i

s_y = standard deviation of dependent variable y

B. DATA TRANSFORMATIONS

The land use and demographic data collected in the field program (and discussed in Section V) were loaded into a field data file on our computer system for processing. Computations were performed on these data to create the model variables chosen for path analysis. These data transformations are summarized in a list of variable definitions in Table 6-1, along with the computer card format used for input. The data form numbers refer to the numbers on the data collection forms, pages 5-9 to 5-20. Certain intermediate calculations on the raw field data are described in footnotes on Table 6-1.

The transformations were performed on data from all 40 field cases and the results stored in two separate data files, one for the Residential model, and one for the Industrial/Office model. Complete listings of all data files are given in Appendix C (published in the separate Appendix to this volume, APTIC Document #80998 [42]). These path analysis data, for both the Residential and Industrial/Office models, were examined for errors by checking variable means and extremes, verifying computer coding sheets and field calculations, and examining the signs of simple correlations between the variables. Errors were then corrected in the field data file and the path variables recomputed.

Analyses were subsequently performed on model variables to test for multicollinearity, possible suppressor variable problems, and the suitability of instrumental variables (used in solving feedback loops). The model was trimmed as a result of the investigations, and the first path analysis performed.

TABLE 6-1

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
1. RES = residential land use in area of influence in 1970 (excluding major project) in dwelling units			
RES = du70t - mpr70			
where: du70t = dwelling units in area of influence in 1970	*	5	49-53
mpr70 = residential land use in major project in 1970 in dwelling units	8	1	36-40
2. COMM = commercial land use in area of influence in 1970 in 1,000 square feet			
COMM = (comm1 + comm2 + comm3 + comm4)/10			
where: comm1 = 100 square feet commercial in area of influence in 1970 (<25K)	29	2	24-28
comm2 = 100 square feet commercial in area of influence in 1970 (25-50K)	30	2	29-33
comm3 = 100 square feet commercial in area of influence in 1970 (50-100K)	31	2	34-38
comm4 = 100 square feet commercial in area of influence in 1970 (>100K)	32	2	39-41

---

* The sum of dwelling units from the block data (page 5-16) plus dwelling units in unblocked areas (page 5-17)

TABLE 6-1 (CONTINUED)

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
3. OFFICE = office land use in area of influence (excluding major project) in 1970 in 1,000 square feet			
OFFICE = (off1 + off2 + off3)/10			
where: off1 = 100 square feet office in area of influence (excluding major project) in 1970 (<50K)	37	2	64-68
off2 = 100 square feet office in area of influence (excluding major project) in 1970 (50-100K)	38	2	69-73
off3 = 100 square feet office in area of influence (excluding major project) in 1970 (>100K)	39	2	74-78
4. MANF = manufacturing land use in area of influence (excluding major project) in 1970 in 1,000 square feet	46	3	34-39
MANF = manf/10			
where: manf = 100 square feet manufacturing in area of influence (excluding major project) in 1970			
5. WHOLE = wholesale/warehouse land use in area of influence in 1970 in 1,000 square feet			
WHOLE = whole/10			
where: whole = 100 square feet wholesale/warehouse in area of influence in 1970	47	3	34-39

TABLE 6-1 (CONTINUED)

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
6. HOTEL = hotel and motel land use in area of influence in 1970 in 1,000 square feet			
HOTEL = (hotel1 + hotel2 + hotel3 + hotel4)/10			
where: hotel1 = 100 square feet hotel in area of influence in 1970 (<25K)	33	2	44-48
hotel2 = 100 square feet hotel in area of influence in 1970 (25-50K)	34	2	49-53
hotel3 = 100 square feet hotel in area of influence in 1970 (50-100K)	35	2	54-58
hotel4 = 100 square feet hotel in area of influence in 1970 (>100K)	36	2	59-61
7. HOSPTL = hospital, etc. land use in area of influence in 1970 in 1,000 square feet			
HOSPTL = (hosp1 +hosp2 + hosp3)/10			
where: hosp1 = 100 square feet hospitals in area of influence in 1970 (25-50K)	40	3	4-8
hosp2 = 100 square feet hospitals in area of influence in 1970 (50-100K)	41	3	9-13
hosp3 = 100 square feet hospitals in area of influence in 1970 (>100K)	42	3	14-18

TABLE 6-1 (CONTINUED)

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
8. CULTUR = cultural land use in area of influence in 1970 in 1,000 square feet			
CULTUR = cultur/10			
where: cultur = 100 square feet cultural in area of influence in 1970	49	3	51-55
9. CHURCH = religious land use in area of influence in 1970 in 1,000 square feet			
CHURCH = church/10			
where: church = 100 square feet religious in area of influence in 1970	48	3	46-50
10. ED = educational land use in area of influence in 1970 in 1,000 square feet			
ED = (ed1 + ed2 +ed3)/10			
where: ed1 = 100 square feet education in area of influence in 1970 (25-50K)	43	3	19-23
ed2 = 100 square feet educational in area of influence in 1970 (25-50K)	44	3	24-28
ed3 = 100 square feet educational in area of influence in 1970 (>100K)	45	3	29-33
11. REC = active outdoor recreational land use in area of influence in 1970 in acres	50	3	57-60
12. HWLM = highway lane miles in area of influence in 1970	74	4	4-8

TABLE 6-1 (CONTINUED)

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
12a. HWLMNX = highway lane miles in area of influence in 1970 without expressways	*	4	9-13
13. MPRT2 = residential land use in major project in year t+2 in dwelling units	5	1	21-25
13a. MPR68 = residential land use in major project in 1968 in dwelling units	7	1	31-35
13b. MPR70 = residential land use in major project in dwelling units	8	1	36-40
14. DUACRE = dwelling units per acre in census tracts in 1960			
DUACRE = (du60c - mpr60)/ac60c			
where: du60c = dwelling units in census tracts	134	7	54-58
ac60c = census tract acreage in 1960	113	8	4-9
mpr60 = dwelling units in major project in 1960	6	1	26-30

* This was computed by adjusting the highway lane miles shown on the data collection form (page 5-12) by the amount of expressway highway lane miles (page 5-13)

TABLE 6-1 (CONTINUED)

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
15. VACACR* = percent vacant developable acreage in area of influence in year (t+0)			
VACACR = vacdev/(10,000-vacund)			
VACUND = vacant undevelopable acreage in area of influence in year (t+0)	82	4	33-37
where: vacdev = vacant developable acreage in area of influence in year (t+0)	83	4	38-41
16. VACHSG = percent vacant housing in census tracts in 1960			
VACHSG = vac60c/du60c			
where: vac60c = vacant available housing units in census tracts in 1960	137	7	69-73
17. HWYINT = highway interchanges in area of influence in year (t+0)	75	4	14-15
18. MINCC = median income factor for families and individuals in census tracts relative to average U.S. income in 1960			
MINCC = mincc/$5,650			
where: mincc = median income for families and individuals	129	7	29-33

* Redefined later, see Table 6-5.

TABLE 6-1 (CONTINUED)

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
19. INCMP = variable indicating the median income level of major project compared to surrounding community in year (t+2)			
INCMP = incmpa - incmpb			
where: incmpa = +1 if major project median income $\geq$ 15% above that of surrounding community	17	1	61
incmpb = +1 if major project median income $\geq$ 15% below that of surrounding community	19	1	63
20. OFFVAC = percent office buildings vacant in metropolitan area in year (t+0)	99	5	35-38
21. OFFACR = office employment per acre in area of influence in year (t+0)			
OFFACR = offemp/10,000			
where: offemp = office employment in area of influence in year (t+0)	95*	4	69-73
22. DISCBD = distance from center of major project to CBD in year (t+0) in miles	79	4	25-28

* Employment data for area of influence (if not available directly) was apportioned from muincipal data by area (page 5-12)

TABLE 6-1 (CONTINUED)

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
23. ENERGY = cost factor for electricity ($/1500 KWH) for users in the metropolitan area in year (t+0) relative to the average U.S. commercial rate in 1960.			
ENERGY = energy/$51.59			
where: energy = dollars per 1500 KWH for commercial users in metropolitan area in year (t+0)	98	5	25-29
24. RRMI = railroad mileage in area of influence in year (t+0)	78	4	20-24
25. WWEA = warehouse and wholesale employment per acre in area of influence in year (t+0)			
WWEA = wwemp/10,000			
where: wwemp = warehouse and wholesale employment in area of influence in year (t+0)	94*	5	74-78

* Employment data for area of influence (if not available directly) was apportioned from municipal data by area (page 5-12)

TABLE 6-1 (CONTINUED

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
26. EMPACR = total employment per acre in area of influence in year (t+0)			
EMPACR = totemp/10,000			
where: totemp = total employment in area of influence in year (t+0)	91*	4	59-63
27. NONHSE = nonhousehold population per acre in census tracts in 1960			
NONHSE = (p60c - hp60c)/ac60c			
where: p60c = total population in census tracts in 1960	127	7	17-22
hp60c = household population in census tracts in 1960	128	7	23-38
28. MPKIDS = school-age children per dwelling unit in major project in year (t+2)	20	1	64-66
29. ENRACR = public school enrollment per acre in census tracts in 1960			
ENRACR = pl460c/ac60c			
where: pl460c = population under 14 years of age in census tracts in 1960	130	7	34-38

* Employment data for area of influence (if not available directly) was apportioned from municipal data (page 5-12)

TABLE 6-1 (CONTINUED)

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
30. MANACR = manufacturing employment per acre in area of influence in year (t+0)			
MANACR = manemp/10,000			
where: manemp = manufacturing employment in area of influence in year (t+0)	92*	4	64-68
31. DELPOP = growth factor for total regional population between 1960 and 1970 (county data)			
DELPOP = (p70cty - p60cty)/p60cty			
where: p70cty = total population in county in 1970	125	7	4-10
p60cty = total population in county in 1970	160	8	10-16
32. DELEMP[1] = growth factor for total regional employment between 1960 and 1970 (county data)			
DELEMP = (e70cty - e60cty)/e60cty			
where: e70cty = total employment in county in 1970	96	5	11-17
e60cty = total employment in county in 1960	95	5	4-10

---

* Employment data for area of influence (if not available directly) was apportioned from municipal data (page 5-12)

[1] Redefined later, see Table 6-5.

TABLE 6-1 (CONTINUED)

PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Data Form Number	Card Number	Column Numbers
33. MINCR = median income factor for the region in year (t+0) relative to the average U.S. income in 1960			
MINCR = mincr/$5,660			
where: mincr = median income for the region in the year (t+0) (county data)	164	5	39-43
34. MPET2 = number of employees in major projects in year (t+2)	26	2	9-13
34a. MPE68 = number of employees in major projects in year 1968	27	2	14-18
34b. MPE70 = number of employees in major projects in year 1970	28	2	19-23
35. AUTO = automobile drivers per acre in county in 1960			
AUTO = au60cy/ac60cy			
where: au60cy = automobile drivers in county in 1960	163	8	30-36
ac60cy = county acreage in 1960	97	5	18-24

C. PROBLEMS IN PATH ANALYSIS

A discussion of our approach to some common limitations of the general multiple regression techniques used in path analysis is given in the following sections:

1. Multicollinearity

Multicollinearity refers to the situation in which some or all of the independent variables in a multiple regression equation are very highly correlated. The effect of multicollinearity on path analysis is to cause instability in the path regression coefficients, and in the extreme case, make it impossible to invert the correlation matrix of the independent variables, preventing solution of the equation. Multicollinearity can exhibit itself as a strong simple correlation between two independent variables, or as the situation where one of the variables is a nearly perfect linear function of a set of the other variables. Multicollinearity typically arises with a simple correlation greater than .90 or .95. In the current study, the test for multicollinearity began with an examination of the simple correlations between the independent variables in all of the path analysis equations. Simple correlations between variables in the residential and industrial models that are significant at the ten percent level are shown on Tables 6-2 and 6-3, respectively. A complete listing of the simple correlation matrices is shown in Appendix C (published in the separate appendix to this volume) [42].

The results show that in general the independent variables of each equation are highly uncorrelated with most values falling below $r=0.35$. This will help improve the stability of the final path regression coefficients. In the Residential model data, the possibility of multicollinearity was found in two cases. In the OFFICE and WHOLE equations, the correlation between the independent variables, HWLM and COMM, is 0.83; and between HWLMNX and COMM it is 0.89, both significant at the 0.1 percent level (t-test). In the OFFICE equation, COMM is eventually trimmed, and in the whole equation both of these variables are eventually trimmed, eliminating possible multicollinearity problems.

In the Industrial/Office model data, multicollinearity was found in only one case. In the WHOLE equation, the correlation between the indepen-

Table 6-2 Residential Model Simple Correlations Significant At p<.10

	MPR70	MPR68	MPRT2	RES	COMM	HOSPTL	CULTUR	CHURCH	EDUC	REC	OFFICE	MANF	WHOLE	HOTEL	HMLM	HMLMNX	HWYINT	DISCBD	EMPACR	MANACR	WMEA	OFFACR	NONHSE	DUACRE	ENRACR	AUTO	OFFVAC	VACHSG	VACACR	MINCR	MINCC	MPKIDS	DELPOP	DELEMP	INCMP	ENERGY	RRMI	
MPR70	1.00	.31	.75	-.40					-.32	.40	-.32		-.42													-.32	.64		-.45	-.36	-.61	-.53		-.31		.36		MPR70
MPR68	.31	1.00	.87							.40			-.39													-.42	.54		-.40	-.34	-.46	-.42		-.35				MPR68
MPRT2	.75	.87	1.00	-.35																						-.41	.49		-.32				-.34	-.45	.33	.32		MPRT2
RES				1.00	.75			.66	.62						.67	.76		-.35					.34	.52	.60	.34			.43		.54			.32		-.51	.34	RES
COMM				.75	1.00	.75		.72	.59						.33	.89		-.59	.52	.48	.37	.63		.30	.44						.45					-.35		COMM
HOSPTL						1.00		.52						.43	.42	.50			.44	.42	.31	.49																HOSPTL
CULTUR							1.00	.31						.31																					-.33			CULTUR
CHURCH						.52	.31	1.00	.75						.69	.73		-.43							.36				.34		.51						-.32	CHURCH
EDUC								.75	1.00		.31		.49		.74	.74		-.57											.62		.76	.45						EDUC
REC										1.00				.33													-.45		-.57						.33			REC
OFFICE											1.00	.32	.45		.36	.40		-.50	.63	.60	.62	.76					-.38				.44					-.35		OFFICE
MANF											.32	1.00	.49	.35				-.36	.30	.33	.30												-.32					MANF
WHOLE											.45	.49	1.00				.51	-.47					.34						.43		.57							WHOLE
HOTEL												.35		1.00												.42						.31						HOTEL
HMLM															1.00	.92		-.61				.39			.44				.43		.62					-.39		HMLM
HMLMNX															.92	1.00		-.60				.41			.45				.44		.50							HMLMNX
HWYINT																	1.00	-.41					.80	.63	.59					.31	.39		-.39					HWYINT
DISCBD																		1.00	-.46	-.45	-.42	-.48			-.30				-.43		-.42			.31		.31		DISCBD
EMPACR																			1.00	.99	.98	.92														-.33		EMPACR
MANACR																			.99	1.00	.98	.88																MANACR
WMEA																			.98	.93	1.00	.87														-.31		WMEA
OFFACR																			.92	.88	.87	1.00					-.40				.32					-.45		OFFACR
NONHSE																							1.00	.71	.63	.36				.41							.48	NONHSE
DUACRE																							.71	1.00	.97													DUACRE
ENRACR																							.63	.97	1.00						.33							ENRACR
AUTO																							.36			1.00			.37	.58	.37	.46		.91	-.39			AUTO
OFFVAC																											1.00		-.36	-.45	-.58	-.43		-.32				OFFVAC
VACHSG																												1.00	-.33	-.49		-.43						VACHSG
VACACR																											-.36	-.33	1.00	.37	.62	.47		.34	-.42			VACACR
MINCR																														1.00	.37	.35	.42	.61		-.39		MINCR
MINCC																														.37	1.00	.52						MINCC
MPKIDS																																1.00		.41				MPKIDS
DELPOP																																	1.00			-.32		DELPOP
DELEMP																																		1.00	-.48			DELEMP
INCMP																																			1.00	-.39	.47	INCMP
ENERGY																																				1.00	-.44	ENERGY
RRMI																																					1.00	RRMI

Table 6-3 Industrial-Office Model Simple Correlations Significant At p<.10

	MPE70	MPE68	MPET2	OFFICE	MANF	COMM	WHOLE	HOTEL	RES	HOSPTL	CULTUR	CHURCH	EDUC	REC	HMLM	HMLMNX	HWYINT	DISCBD	EMPACR	MANACR	WMEA	OFFACR	NONHSE	DUACRE	ENRACR	AUTO	OFFVAC	VACHSG	VACACR	RRMI	ENERGY	DELPOP	DELEMP	MINCR	MINCC	
MPE70	1.00	.83	.33																									.36								MPE70
MPE68	.83	1.00							.35					.43						.36	.32						.34				-.35	.50				MPE68
MPET2	.33		1.00										.52					.43								.32						-.33				MPET2
OFFICE				1.00	.56	.50		.48	.35	.52		.54	.43								.33		.41							.48			.56			OFFICE
MANF				.56	1.00	.73		.68	.31	.56		.47	.43		.50				.38		.58		.50				-.34		.34						.66	MANF
COMM				.50	.73	1.00	.57	.79	.65	.53	.37	.55	.40		.52	.43		-.50	.45	.43	.83		.66	.61	.63				.34							COMM
WHOLE						.57	1.00		.74				.31		.64	.68		-.32	.62	.53	.69	.44	.77	.61	.66											WHOLE
HOTEL				.48	.68	.79		1.00	.38	.50	.51	.68	.31		.38	.68	.31	-.38	.31		.56		.35							.31	-.33	.32				HOTEL
RES									1.00	.32		.34	.48		.72	.64		-.35	.52	.56	.72		.54	.60	.68				.43	.35		.36		.57		RES
HOSPTL									.32	1.00		.67						-.35	.46	.50	.31		.49	.41	.36					.36				-.34	.43	HOSPTL
CULTUR											1.00	.35																			-.37					CULTUR
CHURCH									.34	.67	.35	1.00								.35	.37								.30	.40						CHURCH
EDUC									.48				1.00		.43			.35			.33		.34						.40							EDUC
REC														1.00				.35				.44					.52					.33		.57	.39	REC
HMLM															1.00	.77			.47		.53		.55	.33	.39				.59			.38	.47			HMLM
HMLMNX															.77	1.00	-.45		.55	.50	.56		.46	.42	.48				.51				.32			HMLMNX
HWYINT																-.45	1.00											.42			-.51	.52	.47			HWYINT
DISCBD																		1.00	-.38	-.41	-.36		-.38	-.57	-.55	.64				-.37	.47		.31	.49	.36	DISCBD
EMPACR																			1.00	.83	.70	.71	.50	.57	.54								.40			EMPACR
MANACR																			.88	1.00	.70	.46	.34	.67	.64										-.33	MANACR
WMEA																			.70	.70	1.00	.38	.58	.69	.69											WMEA
OFFACR																			.71	.46	.38	1.00											.70			OFFACR
NONHSE																			.50	.34	.58		1.00	.63	.62	.40				.40			.30			NONHSE
DUACRE																								1.00	.97	-.33					-.30					DUACRE
ENRACR																								.97	1.00											ENRACR
AUTO																								-.33		1.00							.75	.52	.46	AUTO
OFFVAC																											1.00			-.37				.40		OFFVAC
VACHSG																												1.00	.32	.42	.59	.52				VACHSG
VACACR																												.32	1.00			.44	.31	.39		VACACR
RRMI																														1.00				-.35		RRMI
ENERGY																															1.00	-.64				ENERGY
DELPOP																																1.00				DELPOP
DELEMP																																	1.00	.57	.30	DELEMP
MINCR																																		1.00	.75	MINCR
MINCC																																		.75	1.00	MINCC

dent variables WWEA and COMM is 0.83, significant at the 0.1 percent level. The variable COMM is eventually trimmed from this equation, eliminating possible multicollinearity problems.

If significant simple correlations had been found between many independent variables, the next step in testing for multicollinearity would have been to examine the partial correlations between variables. However, since the variables are in general highly uncorrelated, it is not probable that a problem exists in the partial correlations, and so these were not examined.

2. Suppressor Variables

The suppressor variable problem arises in path analysis when multiple regressions are performed between independent and dependent variables which are truly uncorrelated. The presence of an irrelevant variable can cause the sign and magnitude of the path coefficients to deviate from the values one might expect. The approach used for this problem was to examine the magnitude and sign of the simple correlations between dependent and independent variables in the model equations.

In order to test out the hypothesized path structures, the independent variables in each equation were grouped by the significance of their correlation with the dependent variable. The results of this categorization are shown in Table 6-4 and 6-5 for the Residential and Industrial projects, respectively. The significance level of the t-statistic gives the probability that the associated variables are truly uncorrelated, i.e., population r=0. Variables with a significance level above 20 percent generally had correlations below r=0.2. Thus, it can be argued that such variables are essentially uncorrelated with the dependent variable and their inclusion in the model may cause suppressor variable problems. Use of this criterion would trim over half the paths from the current model. The sign of the correlation coefficient associated with each independent variable is indicated in Tables 6-4 and 6-5 by an underline. An examination of the the underlined variables reveals some significant simple correlations that are unexpectedly negative.

Based on these results, it appears a suppressor variable problem could exist with the path model as originally specified. However, it

TABLE 6-4

ANALYSIS OF SIMPLE CORRELATIONS BETWEEN VARIABLES IN THE RESIDENTIAL PATH MODEL*

Dependent Variable	Significance Level of t-Statistic of Correlation with Independent Variables		
	≤ 10%	>10% and ≤20%	>20%
RES	HWLM, HWLMNX, MPRT2, MPR70, DUACRE	MPR68, VACACR, DELPOP	OFFICE, MANF, VACHSG
COMM	RES, OFFICE, HWLM, HWLMNX, MINCC	MANF	MPRT2, MPR68, MPR70, HWYINT, INCMP, DELEMP, MINCR
OFFICE	COMM, MANF, HWLM, HWLMNX, MPR70, OFFVAC, OFFACR, DISCBD	MPRT2, MPR68, VACACR	RES, HWYINT, DELEMP
MANF	MANACR	HWLM	DELEMP, HWLMNX, MPRT2, MPR68, MPR70, HWYINT, ENERGY, RRMI
WHOLE	---	HWLM	COMM, MANF, HWLMNX, HWYINT, RRMI, WWEA, DELEMP
HOTEL	MANF	MPR68	RES, OFFICE, MPRT2, MPR70, DISCBD, EMPACR
HOSPTL	RES	MPR68, MPR70, DISCBD	MPRT2, NONHSE
CULTUR	---	MPR68	MPRT2, MPR70, RES, DISCBD
EDUC	RES, MPR70, MPKIDS	ENRACR	MPRT2, MPR68
REC	MPRT2, MPR68, MPR70, IBCMP	RES, MINCC	---
HWLM	RES, OFFICE, DISCBD, COMM	MPR70, MANF, HWYINT, EMPACR	MPRT2, MPR68, AUTO
HWLMNX	RES, COMM, OFFICE, DISCBD	EMPACR	MANF, MPRT2, MPR68, MPR70, HWYINT, AUTO

---

*An underline signifies a negative correlation

TABLE 6-5

ANALYSIS OF SIMPLE CORRELATIONS BETWEEN VARIABLES IN THE INDUSTRIAL/OFFICE MODEL*

Dependent Variable	Significance of t-Statistic of Correlation with Independent Variables		
	≤ 10%	10% and ≤20%	> 20%
RES	HWLM, HWLMNX, OFFICE, MPE68, DUACRE, $\underline{\text{VACACR}}$, DELPOP, MANF	---	$\underline{\text{MPET2}}$, MPE70, $\underline{\text{VACHSG}}$
COMM	OFFICE, HWLM, HWLMNX, RES, MANF	$\underline{\text{MINCC}}$	MINCR, $\underline{\text{MPET2}}$, MPE68, MPE70, HWYINT, DELEMP
OFFICE	COMM, RES, MANF	$\underline{\text{DISCBD}}$, HWLM, HWLMNX, VACACR	MPET2, $\underline{\text{MPE68}}$, $\underline{\text{MPE70}}$, HWYINT, $\underline{\text{OFFVAC}}$, $\underline{\text{OFFACR}}$, DELEMP
MANF	HWLM, HWLMNX, DELEMP	MPET2, ENERGY, RRMI, MANACR	$\underline{\text{MPE68}}$, $\underline{\text{MPE70}}$, HWYINT
WHOLE	HWLM, HWLMNX, WWEA, COMM	MANF, DELEMP	$\underline{\text{HWYINT}}$, RRMI
HOTEL	OFFICE, $\underline{\text{DISCBD}}$, RES, MANF, EMPACR	MPE68, MPE70	$\underline{\text{MPET2}}$
HOSPTL	RES, $\underline{\text{DISCBD}}$, NONHSE	$\underline{\text{MPE68}}$, $\underline{\text{MPE70}}$	$\underline{\text{MPET2}}$
CULTUR	---	---	RES, $\underline{\text{DISCBD}}$
CHURCH	RES	---	---
EDUC	RES	ENRACR	---
REC	MINCC	---	RES
HWLM	RES, COMM, MANF, EMPACR	MPE68, OFFICE, AUTO	$\underline{\text{MPET2}}$, MPE70, HWYINT, $\underline{\text{DISCBD}}$
HWLMNX	RES, MANF, COMM, $\underline{\text{HWYINT}}$, EMPACR	AUTO, OFFICE	$\underline{\text{MPET2}}$, MPE68, MPE70, $\underline{\text{DISCBD}}$

---

*An underline signifies a negative correlation is not desirable to trim the model using the simple correlations as the only guide. Intercorrelations between independent variables (although not large) can create a "masking" effect whereby an independent variable appears to have a low simple correlation with the dependent variable, but is highly correlated when the effects of other variables are taken into account. Therefore, all trimming that was done to remove possible suppressor variable problems (viz., where path coefficients were insignificant or of the wrong sign), was based on the results of the ordinary and two stage least square multiple regressions. A discussion of the trimming of the model that was performed is given in sections VI.C.4, "Instrumental Variables" and VI.E, "Theory Trimming".

3. Multiple Variable Definitions

There were some independent variables in the system of equations for which we initially had two or three different possible definitions. HWLM and HWLMNX are two representations of highway lane miles. MPRT2, MPR68, and MPR70 are all Major Project (Residential) dwelling units, but in different years. Likewise, MPET2, MPE68, and MPE70 are all Major Project (Industrial) employment. Due to the limited sample size and possible multicollinearity, it is undesirable that all three variables be included in the regression.

In the Residential and Industrial/Office models, HWLM and HWLMNX are about equivalent in the significance of their correlations (see Tables 6-4 and 6-5). The choice of one variable or the other in this case must also take account of the fact that highway lane miles also serves as a dependent variable in one equation. Logically, the endogenous variables linked to highway lane miles by feedback loops (viz., RES, COMM, OFFICE and MANF) should show a stronger causal relationship with non-expressway highway lane miles (HWLMNX) because of local strip highway development along major arterials. The dependent variable chosen must also be significantly correlated with the instrumental variables in the feedback loop system in order for the two-stage least squares method to be successful in estimating path coefficients. Stepwise multiple regressions were performed between the highway lane mile variables and the total set of 20 instrumental variables for the feedback system. In each case the overall goodness of fit of the regression equation was evaluated by testing the significance of the F ratio. This test indicated whether the data sample being analyzed has been drawn from a population in which the multiple correlation is equal to zero, and that any observed multiple correlation is due solely to sampling fluctuations or noise in the data. The results show F ratios significant at the five percent level were occurring only with the HWLMNX variable. None of the HWLM stepwise regressions are significant. Based on all of the above results, the decision was made to use the HWLMNX variable exclusively throughout the model equations.

In the Residential model, the simple correlations (Table 6-4) show MPR70 to be generally the most significant of the three possible Major Project Residential variables. MPRT2 and MPR68 also appear to be significant in several of the equations. However, their inclusion in the model could cause multicollinearity problems since the simple correlations between them and MPR70 are 0.75 and 0.81, respectively. Thus, MPR70 was used exclusively throughout the Residential model equations.

In the Industrial/Office model, all three Major Project Employment variables appear to be significant. An examination reveals the simple correlations between these variables to be:

$$r_{\text{MPET2, MPE68}} = 0.21$$

$$r_{\text{MPET2, MPE70}} = 0.33$$

$$r_{\text{MPE68, MPE70}} = 0.83$$

Only the combinations of MPET2 and MPE68, or MPET2 and MPE70 can be used without causing multicollinearity problems. In order to choose between MPE68 and MPE70, stepwise regressions of all the equations not connected by feedback loops were performed, forcing in MPE68 and MPE70 as the first variable. A comparison of the overall F ratios for these regressions showed MPE70 produced more significant results. Therefore, MPE70 and MPET2 were used throughout the Industrial model.

4. Instrumental Variables

The efficacy of instrumental variables depends first upon causality. They must be a cause of the endogenous variable they are to estimate but not be caused by it. Secondly, instrumental variables should be a stronger causal influence on the endogenous variables whose equations they appear in, than on other endogenous variables in the simultaneous block. Addressing the second point is very difficult since any apparently significant correlation obtained between an instrumental variable and any "other" endogenous variable can be due solely to the feedback loop relationships. Thus, testing this aspect of an instrumental variable involves examination of the initial two-stage least squares analysis.

As discussed previously, a subset of the 20 instrumental variables originally specified in the path analysis model had to be chosen to allow sufficient degrees of freedom in the first stage regressions. Stepwise regressions of each of the five dependent variables in the feedback loop system (viz., RES, COMM, OFFICE, MANF, and HWLMNX) with the total set of 20 instrumental variables were performed. In each case, six of these instrumental variables were first forced into the regression and then the others were allowed to enter in the order of the significance of their added partial contributions. The six forced variables were MPR70, (or MPE70), DUACRE, MINCC, OFFACR, DISCBD, and MANACR. It was decided that at least one major project variable had to appear in each regression. The other five instrumental variables were obtained by choosing the one instrumental variable from each equation that we believed to be the most important, based upon theoretical induced land use relationships and simple correlations between instrumental and endogenous variables. In a two-stage least squares regression of five interconnected equations, a minimum of five instrumental variables is necessary to avoid the problem of underidentification (see Section III.C).

Although all 20 instrumental variables eventually entered all regressions, a cutoff was made after four new variables had entered (beyond the original six forced). An examination of the R^2 change for the instrumental variables (i.e., the additional variance explained in the regression) indicated that after the fourth new variable entered, insignificant amounts of the variance were explained by entering more variables. Next, for both the Residential and Industrial models, a tabulation was made of how many times an instrumental variable appeared in the first four chosen. Finally, an upper limit of 13 instrumental variables (allowing 20-1-13=6 degrees of freedom for the first stage error term) was set for use in the path model. It was felt that six degrees of freedom were a minimum for meaningful results in the first path model run, whereas further trimming beyond this limit would be premature at this point in the analysis. Beyond the six forced variables, seven more were chosen for both the Residential and Industrial/Office models (see Table 6-6), based upon the number of times they appeared in the first four variables chosen in the stepwise regressions. Figures 6-1 and 6-2 show the path models as originally specified, and Figures 6-3 and 6-4 display the models after the above trimming was performed. The dotted lines indicate the variables that were trimmed.

D. PATH DIAGRAMS

The results of the stepwise and two-stage least squares regressions performed for the initial path analysis are presented in Figures 6-5 and 6-6 for the Residential and Industrial/Office Models, respectively. The results of the final path analysis are presented similarly in Figures 6-7 and 6-8. The path coefficients (β) are shown on each path and the R^2 for each regression is displayed next to the associated dependent variable. For each equation, the effect on the dependent variable of all residual causes can be quantified by the path coefficient for residual causes, defined as:

$$\beta_\varepsilon = (1-R^2)^{1/2}$$

TABLE 6-6

INSTRUMENTAL VARIABLES USED IN THE FIRST PATH ANALYSIS

	Residential Model	Industrial/Office Model
1.	DUACRE	DUACRE
2.	MINCC	MINCC
3.	OFFACR	OFFACR
4.	DISCBD	DISCBD
5.	MANACR	MANACR
6.	MPR70	MPE70
7.	VACACR	VACACR
8.	RRMI	RRMI
9.	OFFVAC	OFFVAC
10.	AUTO	AUTO
11.	HWYINT	HWYINT
12.	DELEMP	DELEMP
13.	EMPACR	MPET2

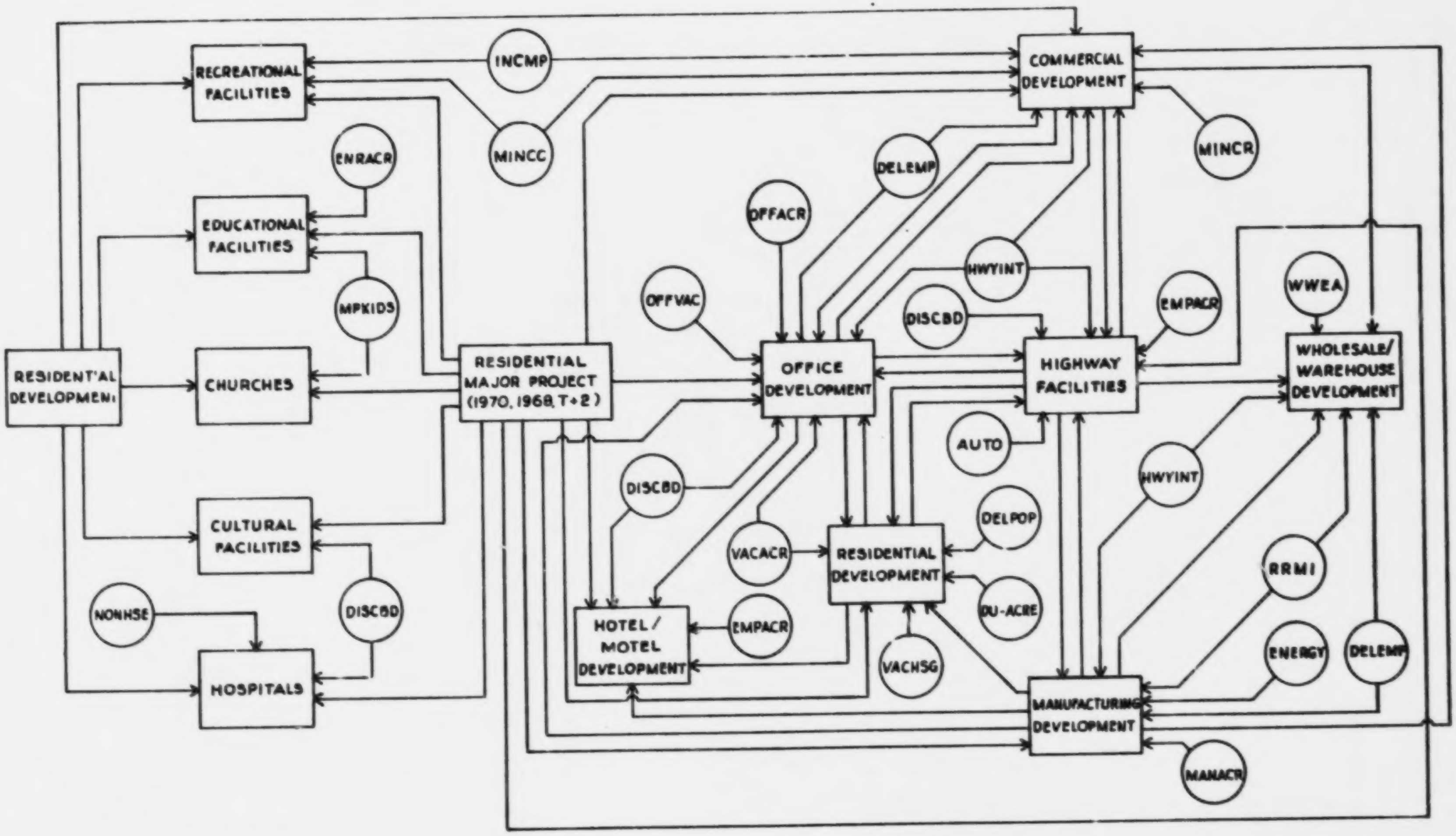

RECREATIONAL FACILITIES
INCMP
COMMERCIAL DEVELOPMENT
MINCC
ENRACR
EDUCATIONAL FACILITIES
DELEMP
OFFACR
MINCR
HWYINT
EMPACR
WWEA
OFFVAC
DISCBD
MPKIDS
RESIDENT'AL DEVELOPMENT
CHURCHES
RESIDENTIAL MAJOR PROJECT (1970, 1968, T+2)
OFFICE DEVELOPMENT
HIGHWAY FACILITIES
WHOLESALE/ WAREHOUSE DEVELOPMENT
AUTO
HWYINT
DISCBD
DELPOP
CULTURAL FACILITIES
VACACR
RESIDENTIAL DEVELOPMENT
RRMI
DISCBD
NONHSE
HOTEL/ MOTEL DEVELOPMENT
EMPACR
DU-ACRE
ENERGY
DELEMP
VACHSG
HOSPITALS
MANUFACTURING DEVELOPMENT
MANACR

Figure 6-1 Original Residential Path Diagram

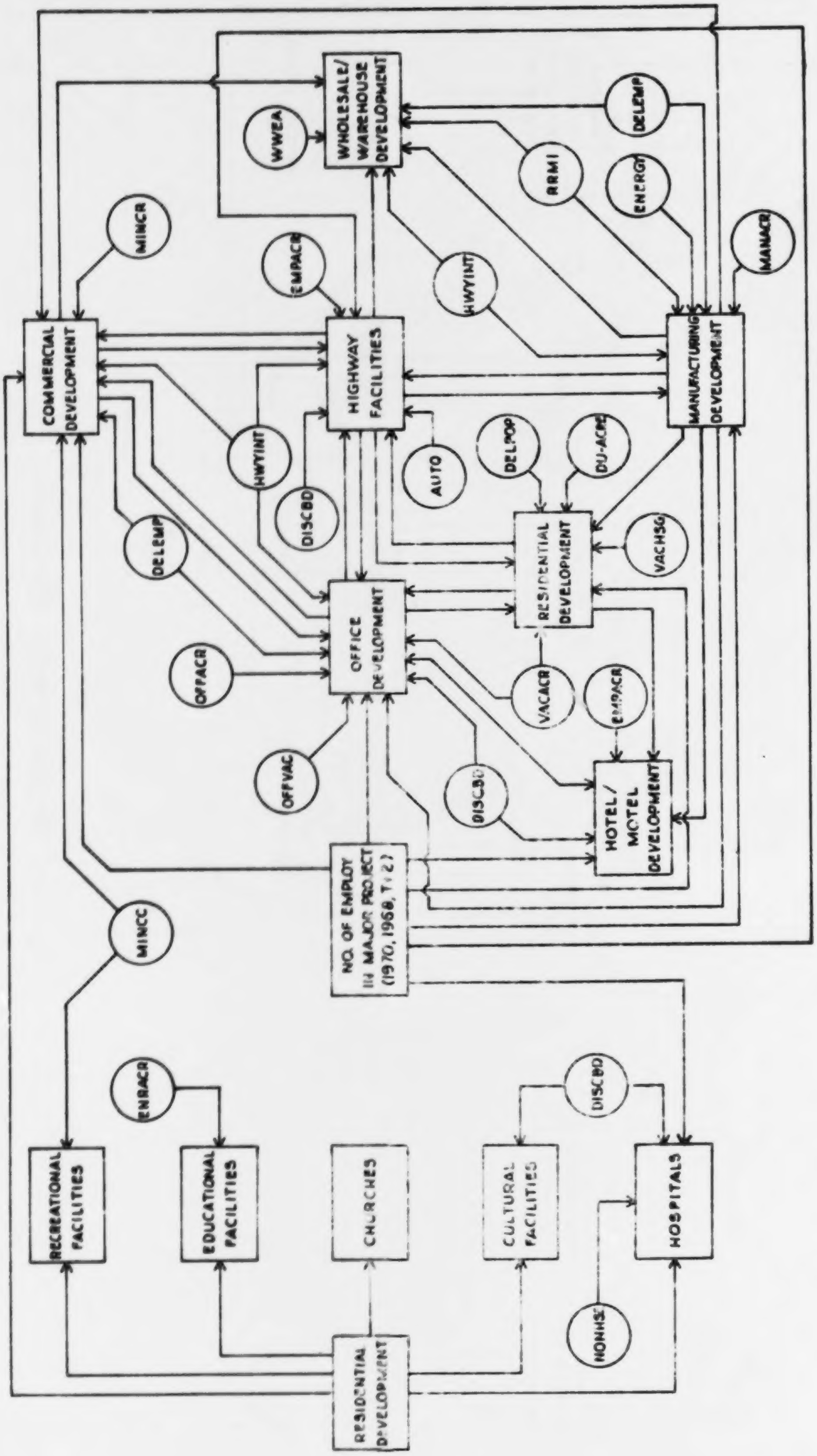

COMMERCIAL DEVELOPMENT
HIGHWAY FACILITIES
WHOLESALE/ WAREHOUSE DEVELOPMENT
MANUFACTURING DEVELOPMENT
RESIDENTIAL DEVELOPMENT
OFFICE DEVELOPMENT
HOTEL/ MOTEL DEVELOPMENT
NO. OF EMPLOY IN MAJOR PROJECT (1970, 1968, T+2)
RECREATIONAL FACILITIES
EDUCATIONAL FACILITIES
CHURCHES
CULTURAL FACILITIES
HOSPITALS
RESIDENTIAL DEVELOPMENT
MINCR
EMPACR
WWEA
RRMI
DELEMP
ENERGY
HWYINT
MANACR
HWYINT
DISCBD
AUTO
DELPOP
DU-ACRE
VACHSG
DELEMP
OFFACR
VACACR
EMPACR
OFFVAC
DISCBD
MINCC
ENRACR
DISCBD
NONHSE

Figure 6-2 Original Industrial-Office Path Diagram

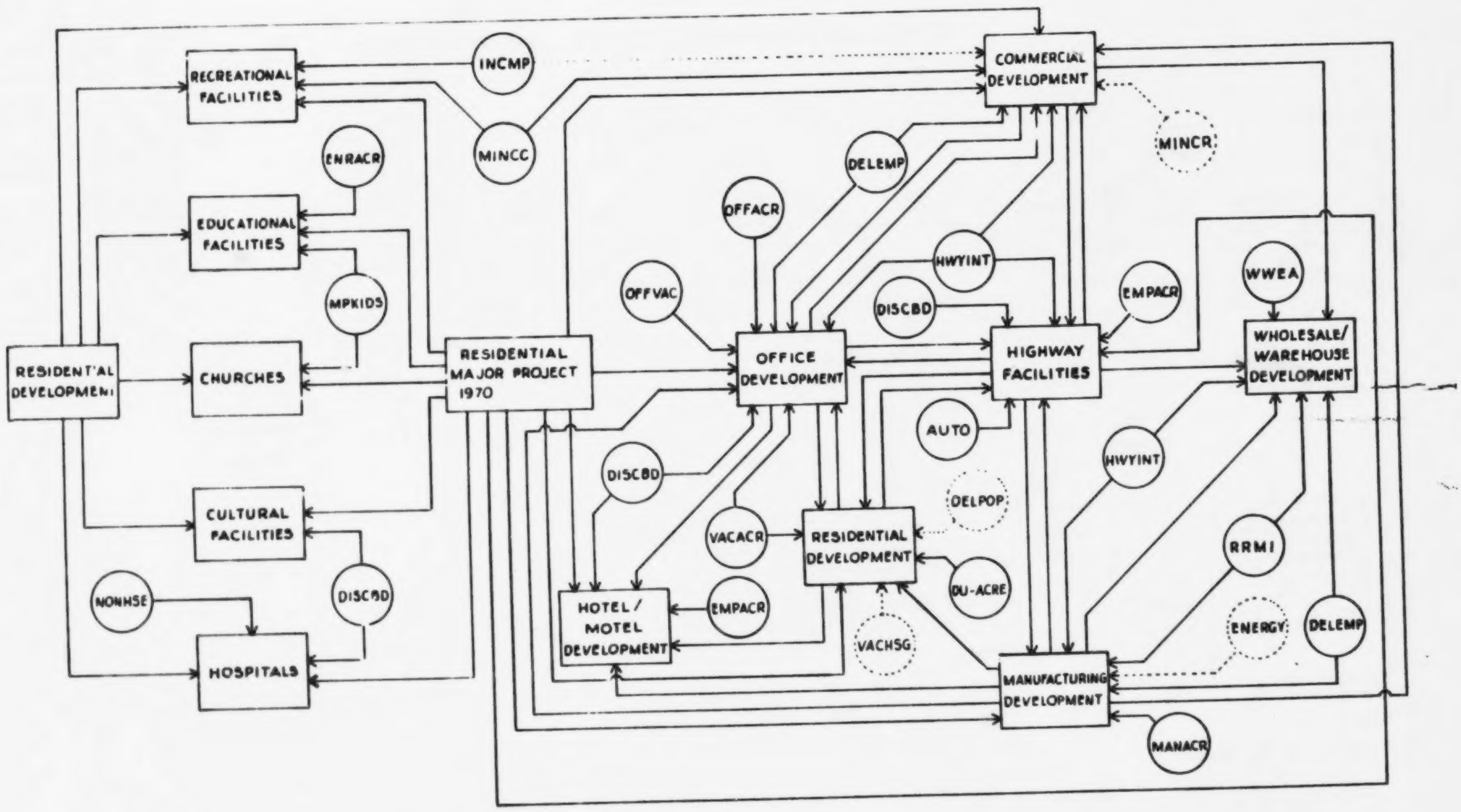

RESIDENT'AL DEVELOPMENI
RECREATIONAL FACILITIES
EDUCATIONAL FACILITIES
CHURCHES
CULTURAL FACILITIES
HOSPITALS
INCMP
MINCC
ENRACR
MPKIDS
DISCBD
NONHSE
RESIDENTIAL MAJOR PROJECT 1970
OFFVAC
OFFACR
DELEMP
OFFICE DEVELOPMENT
DISCBD
VACACR
HOTEL / MOTEL DEVELOPMENT
EMPACR
COMMERCIAL DEVELOPMENT
MINCR
HWYINT
DISCBD
HIGHWAY FACILITIES
EMPACR
AUTO
RESIDENTIAL DEVELOPMENT
DELPOP
DU-ACRE
VACHSG
WWEA
WHOLESALE/ WAREHOUSE DEVELOPMENT
HWYINT
RRMI
ENERGY
DELEMP
MANUFACTURING DEVELOPMENT
MANACR

Figure 6-3 Initial Residential Path Diagram

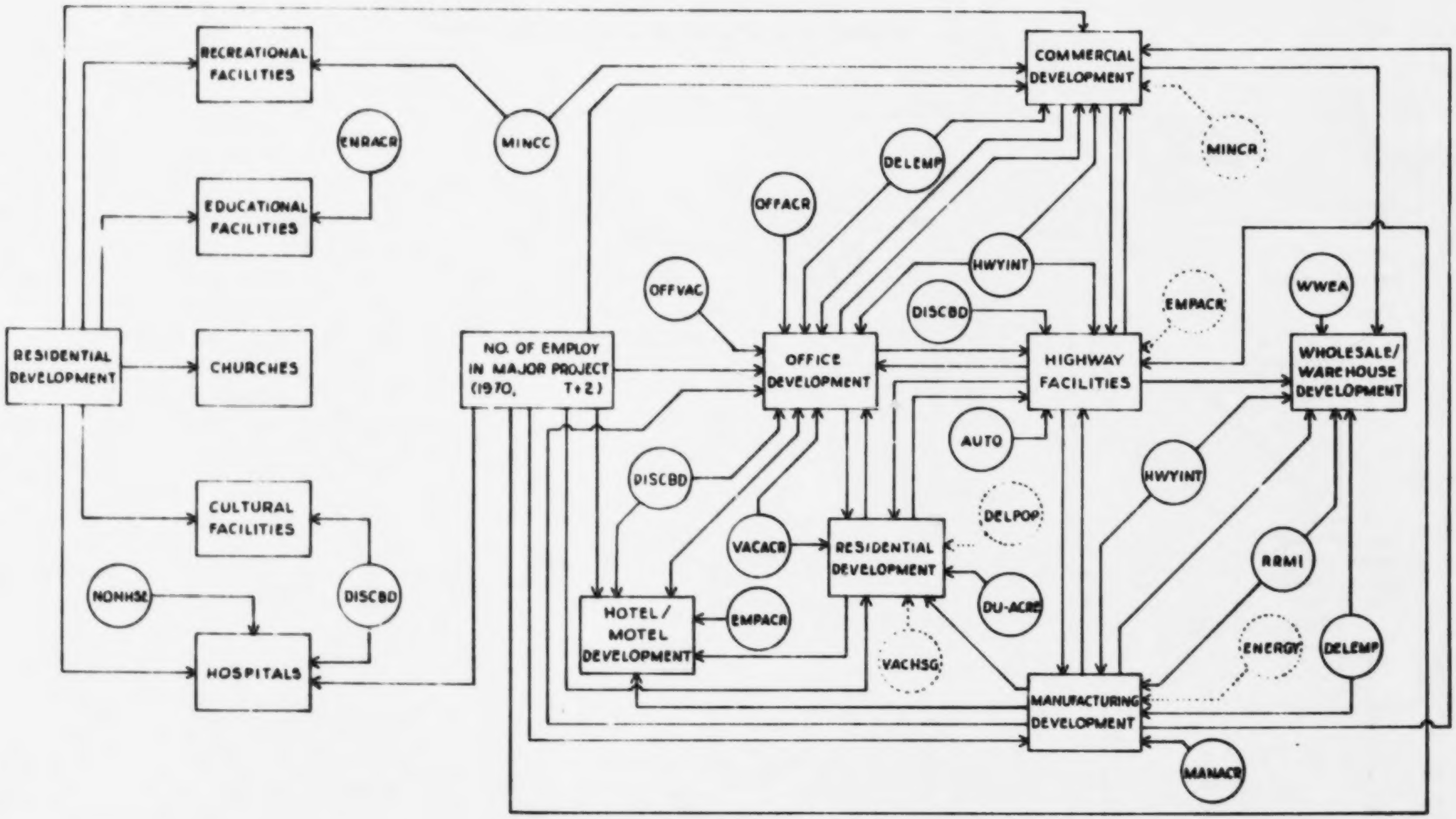

RECREATIONAL FACILITIES
EDUCATIONAL FACILITIES
CHURCHES
CULTURAL FACILITIES
HOSPITALS
RESIDENTIAL DEVELOPMENT
ENRACR
MINCC
DISCBD
NONHSE
NO. OF EMPLOY IN MAJOR PROJECT (1970, T+2)
COMMERCIAL DEVELOPMENT
MINCR
DELEMP
OFFACR
OFFVAC
HWYINT
DISCBD
EMPACR
WWEA
OFFICE DEVELOPMENT
HIGHWAY FACILITIES
WHOLESALE/ WAREHOUSE DEVELOPMENT
AUTO
HWYINT
DISCBD
DELPOP
VACACR
RESIDENTIAL DEVELOPMENT
RRMI
HOTEL/ MOTEL DEVELOPMENT
EMPACR
DU-ACRE
VACHSG
ENERGY
DELEMP
MANUFACTURING DEVELOPMENT
MANACR

Figure 6-4 Initial Industrial-Office Path Diagram

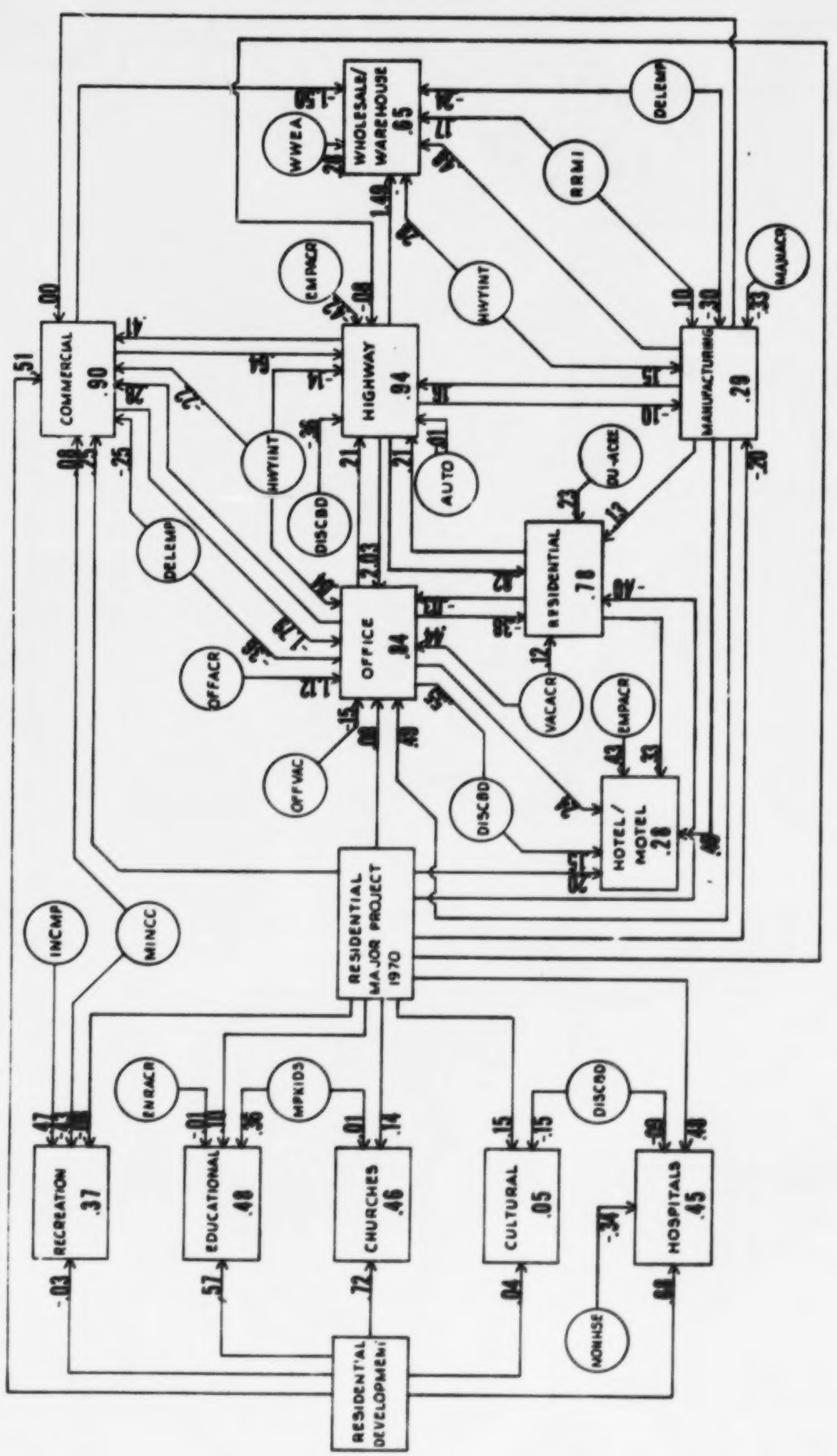

COMMERCIAL .90
HIGHWAY .94
WHOLESALE/ WAREHOUSE .65
MANUFACTURING .29
RESIDENTIAL .78
OFFICE .84
HOTEL/ MOTEL .28
RESIDENTIAL MAJOR PROJECT 1970
RECREATION .37
EDUCATIONAL .48
CHURCHES .46
CULTURAL .05
HOSPITALS .45
RESIDENTIAL DEVELOPMENT
WWEA
DELEMP
RRMI
MAJACR
EMPACR
HWYINT
HWYINT
DISCBD
AUTO
DU-ACRE
DELEMP
OFFACR
VACACR
EMPACR
OFFVAC
DISCBD
INCMP
MINCC
ENRACR
MPKIDS
DISCBD
NONHSE

Figure 6-5 Initial Path Analysis for the Residential Model

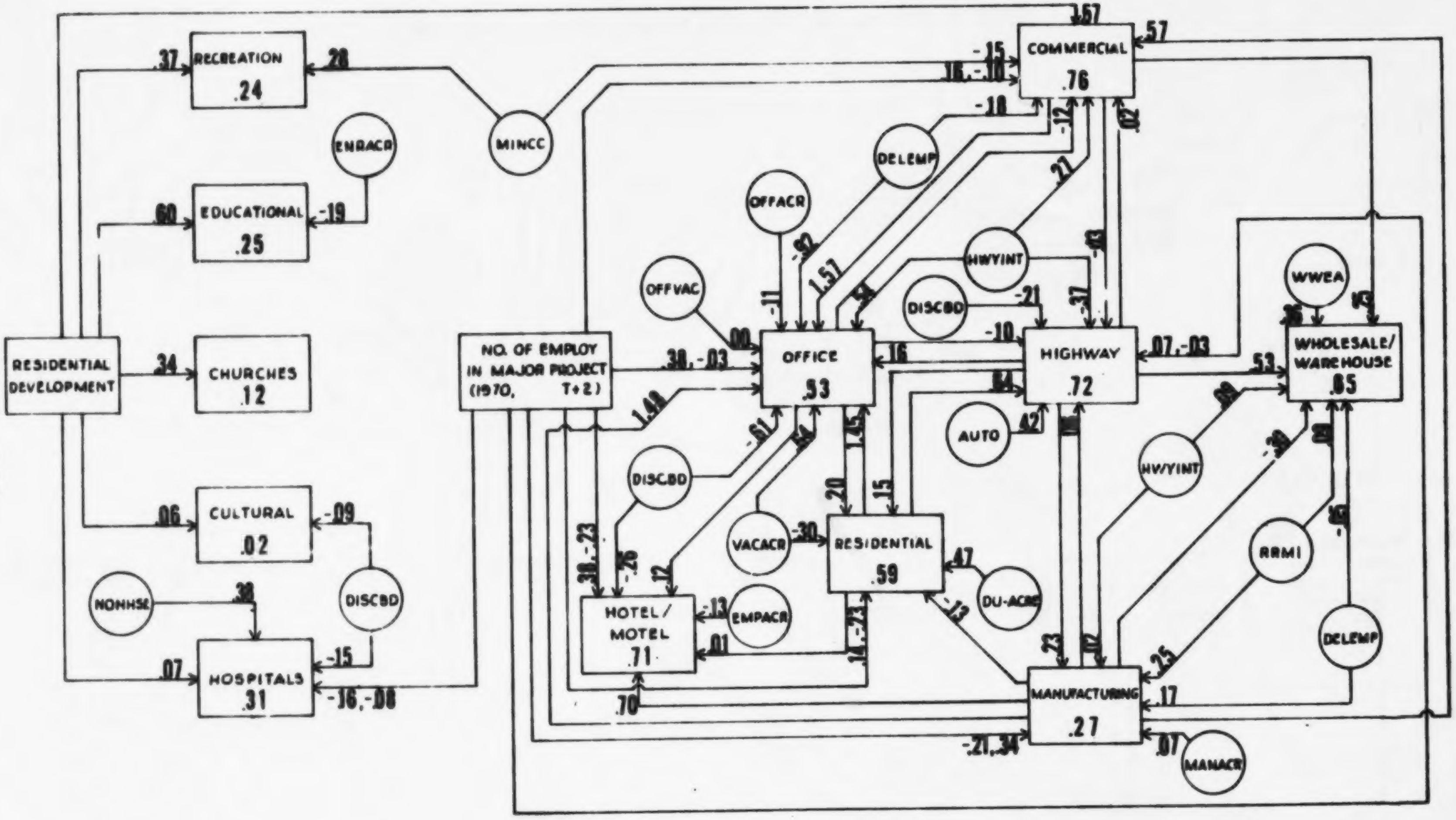

RESIDENTIAL DEVELOPMENT
RECREATION .24
EDUCATIONAL .25
CHURCHES .12
CULTURAL .02
HOSPITALS .31
ENRACR
MINCC
NONHSE
DISCBD
NO. OF EMPLOY IN MAJOR PROJECT (1970, T+2)
OFFICE .53
HIGHWAY .72
COMMERCIAL .76
WHOLESALE/ WAREHOUSE .85
RESIDENTIAL .59
HOTEL/ MOTEL .71
MANUFACTURING .27
OFFVAC
OFFACR
DELEMP
HWYINT
DISCBD
AUTO
VACACR
EMPACR
DU-ACRE
HW/YINT
RRMI
WWEA
MANACR
.37
.28
.60
-.19
.34
.06
-.09
.07
.38
-.15
-.16, -.08
.38, -.03
1.48
.00
-.11
-.92
1.57
.54
-.61
.64
1.45
.16
.20
.15
-.15
.16, -.10
-.10
-.12
.27
.02
.57
.57
-.21
-.37
-.03
-.10
.64
.42
.07, -.03
.53
.36
.19
.69
-.30
.80
.19
.25
.17
.07
.23
.00
-.13
.47
-.30
-.13
.01
.12
-.26
.38, -.23
.14, -.23
.70
-.21, .34

Figure 6-8 Initial Path Analysis for the Industrial-Office Model

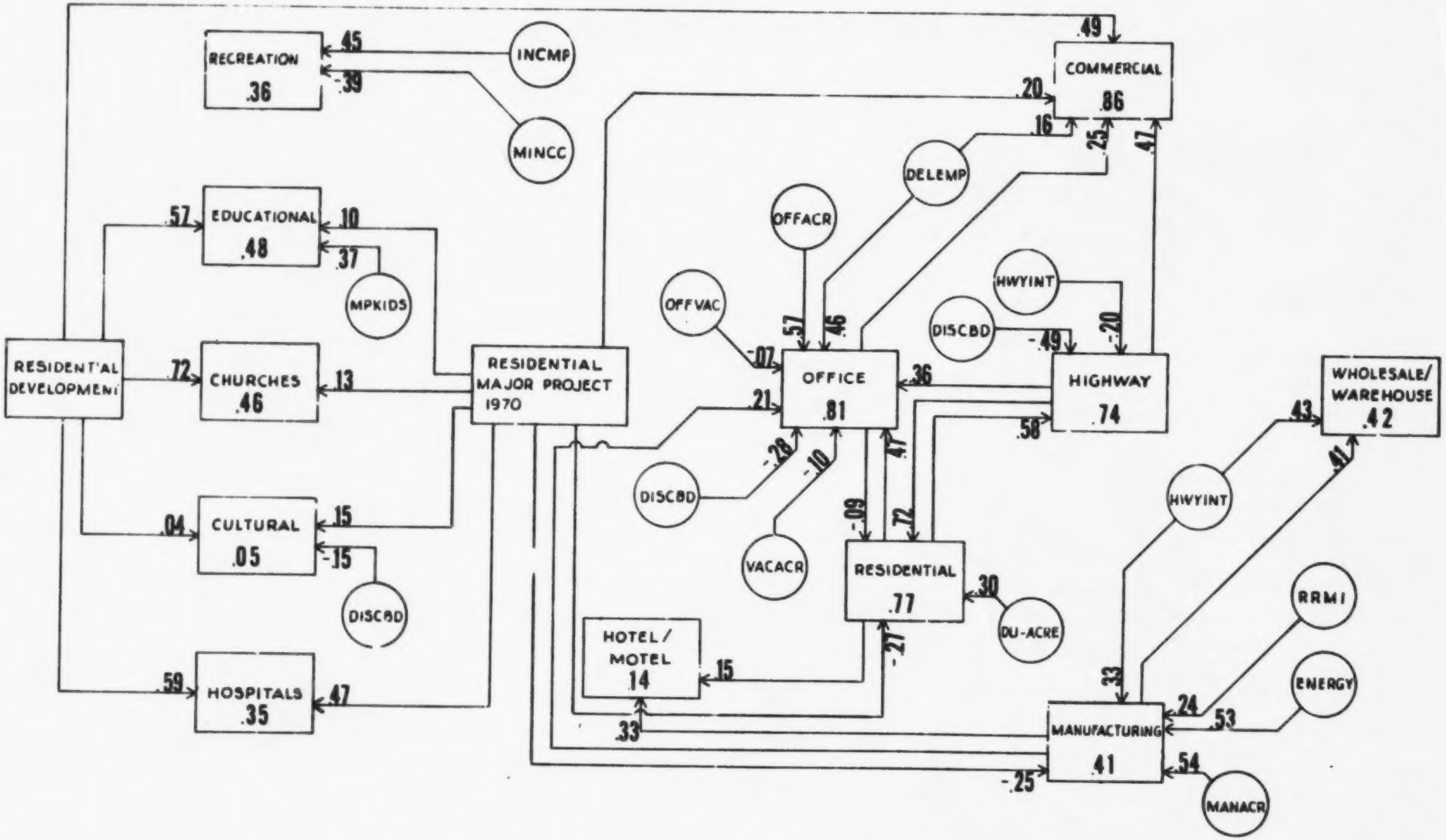

RECREATION .36
EDUCATIONAL .48
CHURCHES .46
CULTURAL .05
HOSPITALS .35
RESIDENT'AL DEVELOPMENT
RESIDENTIAL MAJOR PROJECT 1970
INCMP
MINCC
MPKIDS
DISCBD
OFFVAC
OFFACR
DELEMP
HWYINT
VACACR
DU-ACRE
RRMI
ENERGY
MANACR
COMMERCIAL .86
OFFICE .81
HIGHWAY .74
RESIDENTIAL .77
HOTEL/ MOTEL .14
MANUFACTURING .41
WHOLESALE/ WAREHOUSE .42
.45
-.39
.57
.10
.37
.72
.13
.04
.15
-.15
.59
.47
.49
.20
.16
.25
.47
-.07
.57
.46
.21
-.28
-.10
.36
-.49
-.20
.58
-.09
.47
.72
.30
-.27
.15
.33
.43
.41
.33
.24
.53
.54
-.25

Figure 6-7   Final Path Analysis for the Residential Model

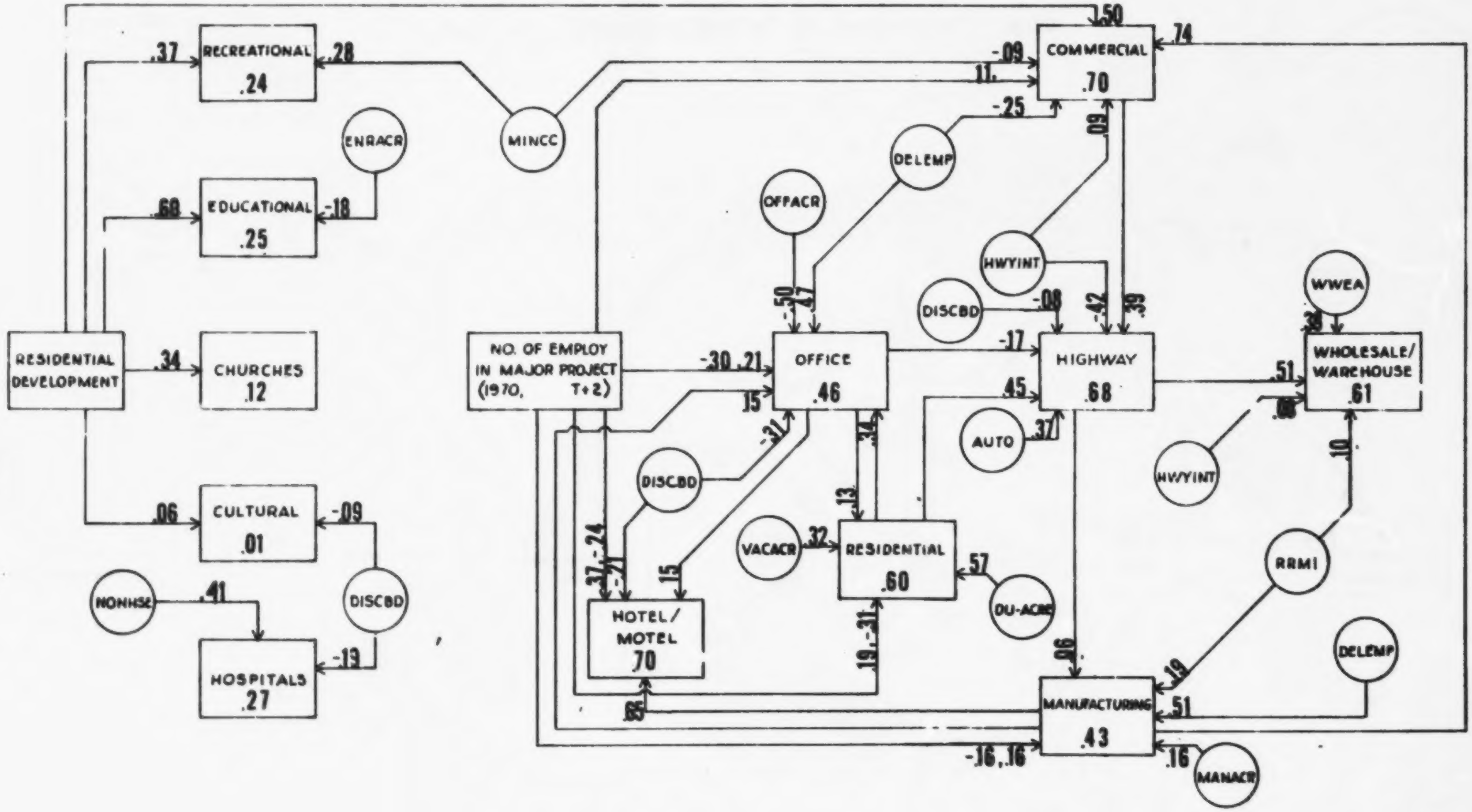

RESIDENTIAL DEVELOPMENT
.37
RECREATIONAL .24
.28
.68
EDUCATIONAL .25
-.18
ENRACR
.34
CHURCHES .12
.06
CULTURAL .01
-.09
DISCBD
NONHSE
.41
HOSPITALS .27
-.19
MINCC
NO. OF EMPLOY IN MAJOR PROJECT (1970, T+2)
-.30 .21
.15
OFFICE .46
OFFACR
-.50
.47
DELEMP
-.25
-.09
.11
.50
COMMERCIAL .70
.74
.60
HWYINT
DISCBD
-.08
-.42
.39
-.17
.45
HIGHWAY .68
AUTO
.37
.51
WHOLESALE/ WAREHOUSE .61
WWEA
.38
.08
HWYINT
.10
RRMI
.19
DISCBD
-.31
.34
.13
VACACR
.32
RESIDENTIAL .60
.57
DU-ACRE
.37 -.24
-.21
.15
HOTEL/ MOTEL .70
.19 -.31
.65
.96
MANUFACTURING .43
.51
DELEMP
-.16 .16
.16
MANACR

Figure 6-8 Final Path Analysis for the Industrial-Office Model

For example, in the final path analysis for the Residential model (Figure 6-7), the R^2 for the RES equation is 0.773. Thus, the path coefficient for all residential causes of RES is $\beta_\varepsilon = \sqrt{1-.773} = 0.48$.

A detailed discussion of the trimming of each model equation is given in the next section, along with a presentation of the path regression coefficients for the final model. The complete statistical output of the final path analysis is given in Appendix C, (published in the separate appendix to this volume, APTIC document 80998) [42].

E. THEORY TRIMMING

1. Approach

The results of the first path analysis revealed many path coefficients that were too low or of the wrong sign, as predicted from theory. In addition, the overall t-statistics of several of the model equations indicated no statistical significance at the five percent level. Thus, criteria were developed, which when applied to the numerical output of the first path analysis, trimmed many model paths. A second path analysis was performed, and the process repeated several times until a final path model was decided upon. This recursive procedure was necessitated by the fact that trimming one variable often causes significant changes in the path coefficients of the remaining variables.

In deciding what paths to drop, we did not choose one fixed significance level (e.g., five percent) against which to compare the model t and F statistics. In the two stage least squares analysis, the values calculated for a t-statistic can only asymptotically be interpreted as being distributed as t, due to the first stage estimation of all feedback loop variables. Thus, it was necessary to use a more general rule in interpreting these statistics. In ordinary least squares analysis, an accepted guideline is that if t or F* is greater than 1.0, the variable being tested raises the adjusted R^2 of the equation, i.e., it explains a significant additional amount of the variance of the dependent variable. Conversely, a t or F less than 1.0 indicates that the variable in question lowers the adjusted R^2 of the equation.

---

* $F(1,N-k-1) = t^2(N-k-1,)$

The adjusted R^2 is defined as:

$$R^2_a = R^2 - \left(\frac{k-1}{N-k}\right)(1-R^2)$$

where:

k = number of independent variables
N = number of data samples.

The conventional R^2 statistic can yield deceptive results when the significance of independent variables is in question. For example, simply adding a variable to any regression equation, whether it is at all correlated with the dependent variable or not, will raise the R^2 and indicate additional variance has been explained. The R^2_a gives a more conservative, unbiased estimate of the amount of variance explained in the dependent variable through the regression equation.

The path coefficient (β) indicates the expected standard deviation change in the dependent variable given a one standard deviation change in a particular independent variable, all other effects held constant. Comparing the β's for different variables allows one to evaluate the relative importance of the various independent variables because of the standardization. In general, $\beta < 0.1$ indicates a variable is not contributing importantly to the regression equation.

In trimming the path analysis model then, the following rules were applied. A path was trimmed if:

- $|t|$ or $F < 1.0$, and $\beta < 0.1$, and loss of the variable would not cause the loss of a significant instrumental variable in the first stage estimations;

- the sign of path coefficient (β) was wrong and counter to the original path model hypothesis.

In addition, some paths whose statistics indicated they should be trimmed were kept if the sign of the path coefficients (β) was correct and the variables were very important causally in the path model. Finally, the discovery of $\beta > 1.0$ in a few model equations indicated correlations between the independent variables to the extent that some paths were redundant. In these instances, simple correlations between the various path coefficients were computed and the most highly correlated variable was trimmed. Correlation coefficients were calculated from the variance-covariance matrix of the estimated coefficients by dividing the covariance of the coefficients of the two variables in question by the product of their individual standard deviations. A detailed discussion of the trimming process performed on each model equation is given in the next section.

2. Discussion of Individual Equations

One result of the first path analysis was that path coefficients for the variables VACACR and DELEMP consistently appeared in the model equations with either an unexpected sign or too low a t or F statistic. Therefore, the original specifications of these variables were examined. It was discovered that VACACR did not exclude vacant developable acreage in the major project, as intended. Also, examination of data for the DELEMP variable indicated that absolute, rather than percentage, employment growth would be a more appropriate predictor of land use development. Therefore, these variables were redefined, as shown on Table 6-7, for the second and subsequent path analyses.

The instrumental variables used for first stage estimates in all analyses of the Industrial/Office model equations are those listed on Table 6-6. For the Residential model, however, the elimination of all feedback loops in the MANF and COMM equations, and theory trimming in the HWLMNX equation, eliminated five instrumental variables (MANACR, RRMI, AUTO, EMPACR, and MINCC) from the list of those used in the first analysis (Table 6-6).

TABLE 6-7

REDEFINITION[1] OF PATH ANALYSIS MODEL DATA TRANSFORMATIONS

Model Variable	Card Number	Column Numbers
15. VACACR = vacant developable acreage in area of influence in 1970, excluding major project		
VACACR = 10,000 - vacund - mpdev - mpund		
where: vacund = vacant undevelopable acreage in area of influence in year (t+0)	4	38-41
mpdev = major project size (acreage)	1	4-8
mpund = undeveloped acreage in area of influence in 1970 under control of the major project developer	1	9-12
32. DELEMP = total regional employment growth per acre between 1960 and 1970 (county data)		
DELEMP = (e70cty - e60cty)/ac60cy		
where: e70cty = total employment in county in 1970	5	11-17
e60cty = total employment in county in 1960	5	4-10
ac60cy = county acreage in 1960	5	18-24

---

[1] See Table 6-1 for original definitions of all variables.

Note that since a path from MANF to OFFICE remains in the final model, MANF had to be added in as an instrumental variable for the two-stage least squares in the Residential model equations. This reduced set of instrumental variables was used for first stage estimates in the second and subsequent analyses of the Residential model equations.

a. Residential Model

(1) RES Equation. The initial path analysis involved a two stage least squares with the following variables:

Endogenous	Exogenous
OFFICE	MPR70
MANF	DUACRE
HWLMNX	VACACR

Based on the second and third analyses, VACACR and MANF were trimmed (low β and t-statistics), respectively. Although the path coefficient for OFFICE subsequently dropped below 0.1 in absolute value, this variable was kept for its causal importance in the model. The final equation became:

$$\mathrm{RES} = -1.38\ \mathrm{OFFICE} + 168\ \mathrm{HWLMNX} - 0.808\ \mathrm{MPR70} + 3930\ \mathrm{DUACRE} + 2730$$

$$\mathit{RES} = -0.09\ \mathit{OFFICE} + 0.72\ \mathit{HWLMNX} - 0.27\ \mathit{MPR70} + 0.30\ \mathit{DUACRE}^{*}$$

$R^2 = 0.773$

$R^2_a = 0.730$

The coefficients for OFFICE and MPR70 are unexpectedly negative. It is possible that these negative signs are due to the finite size of the area of influence. In the case of the first variable it can be argued that office development has priority over and thus uses up land available for residential development. Alternatively, the area-wide changes in zoning preceding office development may preclude nearby residential development. In the case of MPR70, it can be argued that a large residential major project has supplied a good portion of the residential development demand in the area of influence, using up the finite land available. It was hoped that the variable VACACR would control for these problems by representing developable land not in the major project, but unfortunately, it performed poorly in the equation and was dropped.

---

* Italicized variables are standardized and associated coefficients are path coefficients (β).

(2) COMM Equation. The initial path analysis involved a two-stage least squares with the following variables:

Endogenous	Exogenous
RES	MPR70
OFFICE	HWYINT
MANF	MINCC
HWLMNX	DELEMP

Based on the first analysis, MANF was trimmed (low β and t-statistics) and HWYINT was trimmed (wrong sign of β). Subsequently, the loss of the path from COMM to HWLMNX (see section VI.E.2.a.5) removed the only feedback loop in this equation. Thus, the second analysis involved ordinary least squares techniques. Based on the second analysis, MINCC was trimmed (low β and t-statistics) and the final equation became:

$$\begin{aligned} \mathrm{COMM} &= 0.0814\ \mathrm{RES} + 0.649\ \mathrm{OFFICE} + 18.4\ \mathrm{HWLMNX} + 0.0976\ \mathrm{MPR70} \\ &\quad -692\ \mathrm{DELEMP} - 139 \\ \mathit{COMM} &= 0.49\ \mathit{RES} + 0.25\ \mathit{OFFICE} + 0.47\ \mathit{HWLMNX} + 0.20\ \mathit{MPR70} \\ &\quad -0.16\ \mathit{DELEMP} \end{aligned}$$

$R^2 = .860$

$R^2_a = .823$

The coefficient for DELEMP is unexpectedly negative. However, it can be argued it is indicating that areas that experience a large growth in employment from 1960-1970 were less developed areas (those with less total commercial land use in 1970), while older urban areas (those with the greatest total commercial land use in 1970) experienced the lowest growth in employment from 1960-1970.

(3) OFFICE Equation. The initial path analysis involved a two-stage least squares with the following variables:

Endogenous	Exogenous
RES	MPR70
COMM	VACACR
MANF	HWYINT
HWLMNX	OFFVAC
	OFFACR
	DISCBD
	DELEMP

Based on the first analysis, RES and MPR70 were trimmed (low β and t-statistics). Correlations between path coefficients were computed because of the existence of β values > 1.0 for the variables COMM, HWLMNX, and OFFACR. COMM was found to have a correlation with HWLMNX of -0.92, and with OFFACR of 0.78. Based on these results, COMM was trimmed to eliminate a seemingly redundant path. Based on the second analysis, HWLMNX was trimmed (low β and t-statistics) and COMM was added back into the model. Based on the third analysis, HWYINT was trimmed (low β and t-statistics). This analysis also showed COMM to be insignificant (low β and t-statistics), however, it was not trimmed since to do so would have removed the OFFICE equation entirely from the five-equation system of feedback loops. Also, it was not clear whether COMM and HWLMNX were doing so poorly in the model because the feedback loops with RES were lost (in the first trimming) or whether the redefinition of the instrumental variables VACACR and DELEMP affected the first-stage estimates. Therefore, for the fourth analysis, HWLMNX was added back in so that all original endogenous variables, except RES, were present. It was not possible to rerun the original OFFICE equation at this point with all four endogenous variables since the loss of four instrumental variables in the Residential model system (due to trimming of other equations after the first analysis) would have caused the equation to be under-identified and hence unsolvable by two-stage least squares techniques.

The results of the fourth analysis were disappointing with COMM and HWLMNX having path coefficients of β = .09 and -.09, respectively, and low t-statistics. In addition, the β value for MANF was nearly 1.0; indicating a correlation between paths. Correlations between the path coefficients for the three endogenous variables were found to be:

$$r_{MANF,\ COMM} = 0.86$$

$$r_{MANF,\ HWLMNX} = 0.79$$

$$r_{COMM,\ HWLMNX} = 0.98$$

These results indicate a certain degree of redundancy is present between all three endogenous variables, and again, trimming of COMM would eliminate the most redundancy. Thus, in succeeding analyses all combinations of the three endogenous variables, RES, MANF and HWLMNX, were tried. The results indicate that the presence of all three of these endogenous variables allows for the highest R^2_a. Thus, the final equation became:

$$\begin{aligned} OFFICE &= -0.0319\ RES + 5.74\ HWLMNX + 0.0572\ MANF - 0.0127\ VACACR \\ &\quad -5.26\ OFFVAC + 690\ OFFACR - 15.4\ DISCBD + 765\ DELEMP + 421 \\ \mathit{OFFICE} &= -0.47\ \mathit{RES} + 0.36\ \mathit{HWLMNX} + 0.21\ \mathit{MANF} - 0.10\ \mathit{VACACR} \\ &\quad -0.07\ \mathit{OFFVAC} + 0.57\ \mathit{OFFACR} - 0.28\ \mathit{DISCBD} + 0.46\ \mathit{DELEMP} \end{aligned}$$

$R^2 = .812$

$R^2_a = .702$

Unexpected negative signs appear on the path coefficients for RES and VACACR. In the case of RES, the same argument given for the OFFICE to RES path (see Section II.E.2.a.(1)) can be put forth. The negative sign for VACACR is unexplained. This variable was not trimmed, however, because of its importance as an instrumental variable. Although the β value for the variable, OFFVAC, is below 0.1 in absolute value, the sign of the path coefficient is correct. This variable was also not trimmed because of its importance as an instrumental variable.

(4) MANF Equation. The initial path analysis involved a two-stage least squares with the following variables:

Endogenous	Exogenous
HWLMNX	MPR70 HWYINT RRMI MANACR DELEMP

Based on the first analysis, HWLMNX was trimmed (low β and t-statistics), thus, severing the only feedback loop with the simultaneous block of equations in the Residential model. The second and subsequent path analyses, therefore, involved ordinary least squares analysis. Based on the second analysis, RRMI was trimmed (low β and t-statistics). The third analysis indicated a path coefficient for DELEMP that was slightly below 0.1 in absolute value and with an unexpected sign (β=-0.09). The R^2_a for this equation was very low (0.085) and so it was decided to re-analyze the original equation (without HWLMNX) using stepwise multiple regression techniques. In addition to the five exogenous variables shown, the variable, ENERGY, was added into the equation. This variable was trimmed prior to the first path analysis to allow for sufficient degrees of freedom in the first stage estimates of the two-stage least squares (see Section VI.C.4). The use of ordinary least square techniques removed this constraint. Based on the fourth analysis, DELEMP was trimmed (low β and F statistics) and the final equation became:

$$\mathrm{MANF} = 1470\ \mathrm{MANACR} + 316\ \mathrm{HWYINT} + 4100\ \mathrm{ENERGY} - 0.176\ \mathrm{MPR70} + 79.8\ \mathrm{RRMI} - 3490$$

$$MANF = 0.54\ MANACR + 0.33\ HWYINT + 0.53\ ENERGY - 0.25\ MPR70 + 0.24\ RRMI$$

$R^2 = 0.407$
$R^2_a = 0.248$

(5) HWLMNX Equation. The initial path analysis involved a two-stage least squares with the following variables:

Endogenous	Exogenous
RES	MPR70
COMM	HWYINT
OFFICE	DISCBD
MANF	EMPACR
	AUTO

Based on the first analysis, MPR70 and AUTO were trimmed (low β and t-statistics) and EMPACR was trimmed (wrong sign for coefficient and not a significant instrumental variable). Based on the second analysis, OFFICE and MANF were trimmed (low β and t-statistics). Based on the third analysis, COMM was trimmed (low β and t-statistics), and the final equation became:

$$HWLMNX = 0.00247\ RES - 3.31\ HWYINT - 1.73\ DISCBD + 47.9$$
$$\mathit{HWLMNX} = 0.58\ \mathit{RES} - 0.20\ \mathit{HWYINT} - 0.49\ \mathit{DISCBD}$$

$R^2 = 0.742$
$R^2_a = 0.712$

The coefficient for HWYINT is unexpectedly negative. Since HWYINT is an indicator of expressway or highway lane miles, while HWLMNX is defined as non-expressway highway lane miles, it can be argued that areas with an extensive limited access highway network (greatest HWYINT) have less non-expressway highway lane miles, and vice versa.

(6) WHOLE Equation. The initial path analysis involved an ordinary least squares with the following variables:

Endogenous	Exogenous
COMM	DELEMP
MANF	HWYINT
HWLMNX	RRMI
	WWEA

The results of the first analysis revealed both COMM and HWLMNX with β values > 1.0. As a result, HWLMNX was trimmed to eliminate a redundant path. Stepwise multiple regression techniques were used on the second analysis, and based on the results, the variables COMM, DELEMP, RRMI, and WWEA were all trimmed (low β and F statistics). The final equation became:

$$\mathrm{WHOLE} = 60.4\ \mathrm{HWYINT} + 0.0608\ \mathrm{MANF} + 26.5$$
$$\mathit{WHOLE} = 0.43\ \mathit{HWYINT} + 0.41\ \mathit{MANF}$$

$R^2 = 0.421$
$R^2_a = 0.389$

(7) HOTEL Equation. The initial path analysis involved an ordinary least squares with the following variables:

Endogenous	Exogenous
RES	MPR70
OFFICE	DISCBD
MANF	EMPACR

Based on the first analysis, EMPACR was trimmed (wrong sign of β). Stepwise multiple regression techniques were used in the second analysis, and based on the results, the variables, OFFICE, MPR70, and DISCBD, were all trimmed (low β and t-statistics). The final equation became:

$$\mathrm{HOTEL} = 0.00151\ \mathrm{RES} + 0.0140\ \mathrm{MANF} + 49.3$$
$$\mathit{HOTEL} = 0.15\ \mathit{RES} + 0.33\ \mathit{MANF}$$

$R^2 = 0.144$
$R^2_a = 0.096$

(8) HOSPTL Equation. The initial path analysis involved an ordinary least squares with the following variables:

Endogenous	Exogenous
RES	MPR70
	DISCBD
	NONHSE

Based on the first analysis, DISCBD (low β and F statistics) and NONHSE (wrong sign of β) were trimmed. Based on the second analysis, the final equation became:

$$\text{HOSPTL} = 0.0106\ \text{RES} + 0.0246\ \text{MPR70} - 61.6$$
$$\mathit{HOSPTL} = 0.59\ \mathit{RES} + 0.47\ \mathit{MPR70}$$

$R^2 = 0.348$
$R^2_a = 0.312$

(9) CULTUR Equation. The initial path analysis involved an ordinary least squares with the following variables:

Endogenous	Exogenous
RES	MPR70 DISCBD

The results of the first analysis revealed an extremely low R^2_a (actually below zero) indicating the equation is a very poor predictor of cultural facilities. Stepwise multiple regression techniques failed to produce more significant results. Although the β for the variable RES was below 0.1, this variable was not trimmed because it was felt to be an important causal path. The final equation is:

$$\text{CULTUR} = 0.00014\ \text{RES} + 0.00154\ \text{MPR70} - 0.447\ \text{DISCBD} + 19.5$$
$$\mathit{CULTUR} = 0.04\ \mathit{RES} + 0.15\ \mathit{MPR70} - 0.15\ \mathit{DISCBD}$$

$R^2 = 0.050$
$R^2_a = -0.061$

(10) CHURCH Equation. The initial path analysis involved an ordinary least squares with the following variables:

Endogenous	Exogenous
RES	MPR70 MPKIDS

Based on the first analysis, MPKIDS was trimmed (low β and t-statistics); and the final equation became:

$$CHURCH = 0.0134\ RES + 0.00716\ MPR70 + 13.3$$
$$\mathit{CHURCH} = 0.72\ \mathit{RES} + 0.13\ \mathit{MPR70}$$

$R^2 = 0.456$
$R^2_a = 0.425$

(11) EDUC Equation. The initial path analysis involved an ordinary least squares with the following variables:

Endogenous	Exogenous
RES	MPR70
	MPKIDS
	ENRACR

Based on the first analysis, ENRACR was trimmed (wrong sign of β), and the final equation became:

$$EDUC = 0.0392\ RES + 209\ MPKIDS + 0.0203\ MPR70 - 137$$
$$\mathit{EDUC} = 0.57\ \mathit{RES} + 0.37\ \mathit{MPKIDS} + 0.10\ \mathit{MPR70}$$

$R^2 = 0.481$
$R^2_a = 0.420$

(12) REC Equation. The initial path analysis involved an ordinary least squares of the following variables:

Endogenous	Exogenous
RES	INCMP
	MPR70
	MINCC

Based on the first analysis, RES and MPR70 were trimmed (low β and F statistics), and the final equation became:

$$\begin{aligned}
\mathrm{REC} &= 329\ \mathrm{INCMP} - 563\ \mathrm{MINCC} + 826\\
\mathit{REC} &= 0.45\ \mathit{INCMP} - 0.39\ \mathit{MINCC}
\end{aligned}$$

$R^2 = 0.362$
$R^2_a = 0.299$

An unexpected negative sign appears on the coefficient for MINCC. It can be argued, however, that in areas with high income, families have more expensive recreational activities that take them out of their immediate area (e.g., summer house, overseas travel, ski vacations in the mountains, etc.), while in areas of moderate income, there is a greater density of nearby recreational facilities. None of the case studies in the sample included low income areas.

b. Industrial/Office Model

(1) RES Equation. The initial path analysis involved a two stage least squares with the following variables:

Endogenous	Exogenous
OFFICE	MPET2
MANF	MPE70
HWLMNX	DUACRE
	VACACR

Based on the first analysis, MANF was trimmed (wrong sign of β). Based on the second analysis, HWLMNX was trimmed (low β and t-statistics), and the final equation became:

$$\begin{aligned}
\mathrm{RES} &= 3.96\ \mathrm{OFFICE} - 0.859\ \mathrm{MPET2} + 0.392\ \mathrm{MPE70} + 4560\ \mathrm{DUACRE}\\
&\quad + 1.05\ \mathrm{VACACR} - 4860\\
\mathit{RES} &= 0.13\ \mathit{OFFICE} - 0.31\ \mathit{MPET2} + 0.19\ \mathit{MPE70} + 0.57\ \mathit{DUACRE}\\
&\quad + 0.32\ \mathit{VACACR}
\end{aligned}$$

$R^2 = 0.600$
$R^2_a = 0.493$

The one unexpected sign in the path coefficients is for the variable MPET2. First it is important to note that two major project variables (MPET2 and MPE70) appear in the equation. It can be argued that MPE70 is already controlling for the final size of the major project in the equation, and that the coefficient for MPET2 indicates a secondary effect on RES. It is interesting to note that the year t+2 coincides in most cases with the 1960-1961 recession in the U.S., see Table 6-8. Unlike most years, during the period 1960-1961 a large initial growth rate in a major project would not have induced as much of the other types of land use development due to the economic stagnation accompanying a recession. This hypothesis indicates one area where a dynamic, rather than a static, model may have aided in the analysis.

(2) COMM Equation. The initial path analysis involved a two stage least squares with the following variables:

Endogenous	Exogenous
RES	MPET2
OFFICE	MPT70
MANF	HWYINT
HWLMNX	MINCC
	DELEMP

Based on the first analysis, HWLMNX (low β and t-statistics) was trimmed. Although MINCC also had low β and t-statistics, it was not trimmed because of its efficacy as an instrumental variable. Based on the second analysis, OFFICE and MPET2 were trimmed (low β and t-statistics). The results of the third analysis showed the |β| value for HWYINT to be slightly below 0.1, but this variable was not trimmed because it was felt it was important causally in the equation. Thus, the final equation became:

$$\begin{aligned}\mathrm{COMM} &= 0.0785\ \mathrm{RES} + 0.413\ \mathrm{MANF} + 0.0367\ \mathrm{MPE70} + 39.1\ \mathrm{HWYINT}\\ &\quad - 2270\ \mathrm{DELEMP} - 290\ \mathrm{MINCC} - 45.2\\ \mathit{COMM} &= 0.50\ \mathit{RES} + 0.74\ \mathit{MANF} + 0.11\ \mathit{MPE70} + 0.09\ \mathit{HWYINT}\\ &\quad -0.25\ \mathit{DELEMP} - 0.09\ \mathit{MINCC}\end{aligned}$$

$R^2 = 0.794$

$R^2_a = 0.726$

TABLE 6-8

GROSS PRIVATE DOMESTIC INVESTMENT IN THE UNITED STATES, 1955-1973, IN BILLIONS OF 1958 DOLLARS*

Year	Residential Structures	Non-Residential Structures	Total**
1955	25.1	16.2	69.0
1956	22.2	18.5	69.5
1957	20.2	18.2	67.6
1958	20.8	16.6	62.4
1959	24.7	16.2	68.8
1960	21.9	17.4	68.9
1961	21.6	17.4	67.0
1962	23.8	17.9	73.4
1963	24.8	17.9	76.7
1964	24.2	19.1	81.9
1965	23.8	22.3	90.1
1966	21.3	24.0	95.4
1967	20.4	22.6	93.5
1968	23.2	23.4	98.8
1969	23.7	24.3	103.8
1970	22.2	23.7	99.5
1971	29.0	22.5	105.0
1972	34.6	23.0	118.3
1973	34.0	24.8	126.6

* Source: Department of Commerce, Bureau of Economic Analysis

**Includes change in Business Inventories and Fixed Investment in Residential Structures, Non-Residential Structures and Producers Durable Equipment

Besides the previously discussed MPET2, the only variable with an unexpected sign on its coefficient was DELEMP. In this instance, it can be argued that the greatest growth in employment between 1960 and 1970 occurred in largely undeveloped areas which, even in 1970, probably have less total commercial land use than older urban areas (a bias towards low initial levels).

(3) OFFICE Equation. The initial path analysis involved two stage least squares with the following variables:

Endogenous	Exogenous
RES	MPET2
COMM	MPE70
MANF	VACACR
HWLMNX	HWYINT
	OFFVAC
	OFFACR
	DISCBD
	DELEMP

The results of the first analysis revealed $|\beta|$ values > 1.0 for three variables: RES, COMM, and MANF. An examination of the correlations between both regression coefficients indicated COMM was the most highly correlated variable. It was, therefore, trimmed to eliminate a redundant path. Based on the second and third analyses, the variables OFFVAC and HWLMNX were trimmed (wrong sign of β), respectively. Based on the fourth analysis, VACACR and HWYINT were trimmed (low β and t-statistics), and the final equation became:

$$\begin{aligned}\mathrm{OFFICE} = {} & 0.0110\ \mathrm{RES} + 0.0178\ \mathrm{MANF} - 0.0199\ \mathrm{MPE70} + 0.0195\ \mathrm{MPET2} \\ & - 285\ \mathrm{OFFACR} - 8.11\ \mathrm{DISCBD} + 902\ \mathrm{DELEMP} + 174\end{aligned}$$

$$\begin{aligned}\mathit{OFFICE} = {} & 0.34\ \mathit{RES} + 0.15\ \mathit{MANF} - 0.30\ \mathit{MPE70} + 0.21\ \mathit{MPET2} \\ & -0.50\ \mathit{OFFACR} - 0.31\ \mathit{DISCBD} + 0.47\ \mathit{DELEMP}\end{aligned}$$

$R^2 = 0.462$
$R^2_a = 0.214$

Examination of the statistical output for the final equation reveals unexpected signs on the coefficients for the variables OFFACR, MPET2, and MPE70.

OFFACR was not trimmed even though it was felt its coefficient should not be negative because of its importance as an instrumental variable. The signs for MPET2 and MPE70 appear to be reversed from the situation discussed previously in the RES equation. Again we believe that MPE70 is controlling for the final size of the major project in the equation. The negative sign on MPE70 can be explained by the argument that where the Industrial/Office major project is large, the presence of this development discouraged other office development in the same area, using up most of the available land (zoned for office) and capital resources. As before, the variable MPET2 can be thought of as controlling secondary effects. We were unable, however, to formulate a reason why the coefficient was not negative, as in the RES equation.

(4) MANF Equation. The initial path analysis involved two stage least squares of the following variables:

Endogenous	Exogenous
HWLMNX	MPET2 MPE70 HWYINT RRMI MANACR DELEMP

Based on the first analysis, HWYINT (low β and t-statistics) was trimmed. The results of the second analysis indicated low β and t-statistics for HWLMNX, but this variable was kept because of its importance causally in the model. Thus, the final equation became:

$$\begin{aligned} \text{MANF} &= 4.55\ \text{HWLMNX} + 0.124\ \text{MPET2} - 0.0974\ \text{MPE70} + 55.6\ \text{RRMI} \\ &\quad + 8690\ \text{DELEMP} + 1020\ \text{MANACR} - 22.5 \\ \textit{MANF} &= 0.06\ \textit{HWLMNX} + 0.16\ \textit{MPET2} - 0.16\ \textit{MPE70} + 0.19\ \textit{RRMI} \\ &\quad + 0.51\ \textit{DELEMP} + 0.16\ \textit{MANACR} \end{aligned}$$

$R^2 = 0.426$

$R^2_a = 0.221$

Note the unexpected signs on the coefficients for MPET2 and MPE70 indicate the same situation is occurring as was discussed previously for the OFFICE equation.

(5) HWLMNX Equation. The initial path analysis involved two stage least squares of the following variables:

Endogenous	Exogenous
RES	MPET2
COMM	MPE70
OFFICE	HWYINT
MANF	DISCBD
	AUTO

Based on the first, second, and third analyses, MPET2, MANF, and MPE70 were trimmed (low β and t-statistics), respectively. The variable DISCBD was not trimmed as a result of low β and t-statistics in the fourth analysis because of its importance causally in the model. Thus, the final equation became:

$$\begin{aligned} HWLMNX &= 0.00178\ RES - 0.0209\ OFFICE + 0.00980\ COMM - 4.59\ HWYINT \\ &\quad -0.246\ DISCBD + 32.6\ AUTO + 14.3 \\ \mathit{HWLMNX} &= 0.45\ \mathit{RES} - 0.17\ \mathit{OFFICE} + 0.39\ \mathit{COMM} - 0.42\ \mathit{HWYINT} \\ &\quad -0.08\ \mathit{DISCBD} + 0.37\ \mathit{AUTO} \end{aligned}$$

$R^2 = 0.682$
$R^2_a = 0.568$

The coefficient for HWYINT is unexpectedly negative. Since HWYINT is an indicator of the presence or absence of expressway or interstate highways, while HWLMNX is defined as non-expressway highway lane miles, it can be argued that areas with an extensive interstate highway network (greatest HWYINT) have less non-expressway highway lane miles, and vice-versa. The coefficient for OFFICE is also unexpectedly negative. However, a reasonable argument for this variable could not be developed.

(6) WHOLE Equation. The initial path analysis involved ordinary least squares of the following variables:

Endogenous	Exogenous
COMM	HWYINT
MANF	RRMI
HWLMNX	WWEA
	DELEMP

Based on the first analysis, MANF was trimmed (wrong sign of β). Based on the second COMM and DELEMP were trimmed (low β and F-statistics) and the final equation became:

$$\text{WHOLE} = 10{,}500\ \text{WWEA} + 35.7\ \text{HWLMNX} + 29.0\ \text{RRMI} + 64.6\ \text{HWYINT} - 1{,}120$$
$$\mathit{WHOLE}\quad 0.38\ \mathit{WWEA} + 0.51\ \mathit{HWLMNX} + 0.10\ \mathit{RRMI} + 0.08\ \mathit{HWYINT}$$

$R^2 = 0.615$

$R^2_a = 0.543$

The variable HWYINT was not trimmed, despite its low β value, because of its causal importance in the model.

(7) HOTEL Equation. The initial path analysis involved ordinary least squares of the following variables:

Endogenous	Exogenous
RES	MPE70
OFFICE	DISCBD
MANF	EMPACR
	MPET2

Based on the first analysis, RES (low β and F-statistics) and EMPACR (wrong sign of β) were trimmed, and the final equation became:

$$\text{HOTEL} = 0.0524\ \text{MANF} + 0.105\ \text{OFFICE} + 0.0182\ \text{MPE70} - 0.0156\ \text{MPET2} - 4.04\ \text{DISCBD} - 26.0$$
$$\mathit{HOTEL} = 0.65\ \mathit{MANF} + 0.15\ \mathit{OFFICE} + 0.37\ \mathit{MPE70} - 0.24\ \mathit{MPET2} - 0.21\ \mathit{DISCBD}$$

$R^2 = 0.699$

$R^2_a = 0.619$

The opposite signs of the coefficients for MPE70 and MPET2 indicate the situation discussed previously for the RES equation is occurring.

(8) HOSPTL Equation. The initial path analysis involved ordinary least squares of the following variables:

Endogenous	Exogenous
RES	MPE70 DISCBD NONHSE MPET2

Based on the first analysis, MPE70 (wrong sign of β) and MPET2 (low β and F-statistics) were trimmed. Based on the second analysis, RES was trimmed (low β and F-statistics) and the final equation became:

$$\mathrm{HOSPTL} = 508\ \mathrm{NONHSE} - 5.66\ \mathrm{DISCBD} + 108$$
$$HOSPTL = 0.41\ NONHSE - 0.19\ DISCBD$$

$R^2 = 0.269$
$R^2_a = 0.228$

(9) CULTUR Equation. The initial path analysis involved ordinary least squares of the following variables:

Endogenous	Exogenous
RES	DISCBD

The results of the first analysis indicate both of these variables should be trimmed. However, to do so would totally eliminate the equation. Thus, the final form is:

$$\mathrm{CULTUR} = 0.0023\ \mathrm{RES} - 0.249\ \mathrm{DISCBD} + 13.6$$
$$CULTUR = 0.06\ RES - 0.09\ DISCBD$$

$R^2 = 0.015$
$R^2_a = -0.040$

The extremely low R^2 (the adjusted R^2 is actually below zero) indicates that similar to the CULTUR equation in the Residential model, this equation is a very poor predictor of cultural facilities.

(10) CHURCH Equation. The initial path analysis involved ordinary least squares of the following variables.

Endogenous
RES

The final equation is:

CHURCH = 0.00513 RES + 51.1
CHURCH = 0.34 *RES*

$R^2 = 0.119$
$R^2_a = 0.119$

(11) EDUC Equation. The initial path analysis involved ordinary least squares of the following variables:

Endogenous	Exogenous
RES	ENRACR

Based on the first analysis, ENRACR was trimmed (wrong sign of β), and the final equation became:

EDUC = 0.0405 RES + 304
EDUC = 0.48 *RES*

$R^2 = 0.227$
$R^2_a = 0.227$

(12) REC Equation. The initial path analysis involved ordinary least squares of the following variables:

Endogenous	Exogenous
RES	MINCC

The final equation is:

$$\mathrm{REC} = 0.0149\ \mathrm{RES} + 242\ \mathrm{MINCC} - 184$$
$$\mathit{REC} = 0.37\ \mathit{RES} + 0.28\ \mathit{MINCC}$$

$$R^2 = 0.239$$
$$R^2_a = 0.176$$

3. Summary of Final Model Equations in Unstandardized Form

a. English Units

(1) Residential Model

$$\mathrm{RES} = -1.38\ \mathrm{OFFICE} + 168\ \mathrm{HWLMNX} - 0.808\ \mathrm{MPR70} + 3930\ \mathrm{DUACRE} + 2730$$

$$\mathrm{COMM} = 0.0814\ \mathrm{RES} + 0.649\ \mathrm{OFFICE} + 18.4\ \mathrm{HWLMNX} + 0.0976\ \mathrm{MPR70} - 692\ \mathrm{DELEMP} - 139$$

$$\mathrm{OFFICE} = -0.0319\ \mathrm{RES} + 5.74\ \mathrm{HWLMNX} + 0.0572\ \mathrm{MANF} - 0.0127\ \mathrm{VACACR} - 5.26\ \mathrm{OFFVAC} + 690\ \mathrm{OFFACR} - 15.4\ \mathrm{DISCBD} + 765\ \mathrm{DELEMP} + 421$$

$$\mathrm{MANF} = 1470\ \mathrm{MANACR} + 316\ \mathrm{HWYINT} + 4100\ \mathrm{ENERGY} - 0.176\ \mathrm{MPR70} + 79.8\ \mathrm{RRMI} - 3940$$

$$\mathrm{HWLMNX} = 0.00247\ \mathrm{RES} - 3.31\ \mathrm{HWYINT} - 1.73\ \mathrm{DISCBD} + 47.9$$

$$\mathrm{WHOLE} = 60.4\ \mathrm{HWYINT} + 0.0608\ \mathrm{MANF} + 26.5$$

$$\mathrm{HOTEL} = 0.00151\ \mathrm{RES} + 0.0140\ \mathrm{MANF} + 49.3$$

$$\mathrm{HOSPTL} = 0.0106\ \mathrm{RES} + 0.0246\ \mathrm{MPR70} - 61.6$$

$$\mathrm{CULTUR} = 0.00014\ \mathrm{RES} + 0.00154\ \mathrm{MPR70} - 0.447\ \mathrm{DISCBD} + 19.5$$

$$\mathrm{CHURCH} = 0.0134\ \mathrm{RES} + 0.00716\ \mathrm{MPR70} + 13.3$$

$$\mathrm{EDUC} = 0.0392\ \mathrm{RES} + 209\ \mathrm{MPKIDS} + 0.0203\ \mathrm{MPR70} - 137$$

$$\mathrm{REC} = 329\ \mathrm{INCMP} - 563\ \mathrm{MINCC} + 826$$

(2) Industrial/Office Model

$$\mathrm{RES} = 3.96\ \mathrm{OFFICE} - 0.859\ \mathrm{MPET2} + 0.392\ \mathrm{MPE70} + 4560\ \mathrm{DUACRE} + 1.05\ \mathrm{VACACR} - 4860$$

$$\mathrm{COMM} = 0.0785\ \mathrm{RES} + 0.413\ \mathrm{MANF} + 0.0367\ \mathrm{MPE70} + 39.1\ \mathrm{HWYINT} - 2270\ \mathrm{DELEMP} - 290\ \mathrm{MINCC} - 45.2$$

OFFICE = 0.0110 RES + 0.0178 MANF - 0.0199 MPE70 + 0.0195 MPET2 - 285 OFFACR - 8.11 DISCBD + 902 DELEMP + 174

MANF = 4.55 HWLMNX + 0.124 MPET2 - 0.0974 MPE70 + 55.6 RRMI + 8690 DELEMP + 1020 MANACR - 22.5

HWLMNX = 0.00178 RES - 0.0209 OFFICE + 0.00980 COMM - 4.59 HWYINT - 0.246 DISCBD + 32.6 AUTO + 14.3

WHOLE = 10,500 WWEA + 34.7 HWLMNX + 29.0 RRMI + 64.6 HWYINT - 1,120

HOTEL = 0.0524 MANF + 0.105 OFFICE + 0.0182 MPE70 - 0.0156 MPET2 - 4.04 DISCBD - 26.0

HOSPTL = 508 NONHSE - 5.66 DISCBD + 108

CULTUR = 0.0023 RES - 0.249 DISCBD + 13.6

CHURCH = 0.00513 RES + 51.1

EDUC = 0.0405 RES + 304

REC = 0.0149 RES + 242 MINCC - 184

b. Metric Units

As can be seen in Tables 6-1 and 6-5, the model variables were defined and the path analyses carried out in English units. Thus, the path regression coefficients given for each equation in the previous section reflect this fact. The path diagrams in Section VI.D, however, display the path coefficients (β) which are independent of the units chosen.

This section summarizes the final model equations for variables RES, COMM, OFFICE, MANF, WHOLE, HOTEL, HOSPTL, CULTUR, CHURCH, and EDUC. The variables REC, VACACR, also use m^2, but in place of acres. Distances in miles in HWLMNX, DISCBD, and RRMI are converted to km. And finally, the units of acre^{-1} are replaced by m^{-2} in the variables DUACRE, OFFACR, WWEA, EMPACR, NONHSE, ENRACR, MANACR, DELEMP, and AUTO.

The conversion factors used were:

1,000 ft^2 → m^2 *by 92.90
acres → m^2 *by 4,047
miles → km *by 1.609
acre^{-1} → m^{-2} ÷by 4047

(1) Residential Model

RES = - 1.38 OFFICE + 9700 HWLMNX - 75.1 MPR70 + 1,480,000,000 DUACRE
+ 254,000

COMM = 0.0814 RES + 0.649 OFFICE + 1,060 HWLMNX + 9.07 MPR70
- 260,000,000 DELEMP - 12,900

OFFICE = - 0.0319 RES + 331 HWLMNX + 0.0572 MANF - 0.000292 VACACR
- 489 OFFVAC + 260,000,000 OFFACR - 889 DISCBD + 288,000,000 DELEMP
+ 39,100

MANF = 553,000,000 MANACR + 29,400 HWYINT + 381,000 ENERGY - 16.4 MPR70
+ 4,610 RRMI - 366,000

HWLMNX = 0.0000428 RES - 5.33 HWYINT - 1.73 DISCBD + 771.

WHOLE = 5,610 HWYINT + 0.0608 MANF + 2,460

HOTEL = 0.00151 RES + 0.0140 MANF + 4,580

HOSPTL = 0.0106 RES + 2.29 MPR70 - 5,720

CULTUR = 0.00014 RES + 0.143 MPR70 - 25.8 DISCBD + 1,810

CHURCH = 0.0134 RES + 0.665 MPR70 + 1,240

EDUC = 0.0392 RES + 19,400 MPKIDS + 1.89 MPR70 - 12,700

REC = 1,330,000 INCMP - 2,280,000 MINCC + 3,340,000

(2) Industrial/Office Model

RES = 3.96 OFFICE - 79.8 MPET2 + 36.4 MPE70 + 1,710,000,000 DUACRE
+ 0.0241 DUACRE - 451,000

COMM = 0.0785 RES + 0.413 MANF + 3.41 MPE70 + 3,630 HWYINT
- 853,000,000 DELEMP - 26,900 MINCC - 4,200

OFFICE = 0.0110 RES + 0.0178 MANF - 1.85 MPE70 + 1.81 MPET2
- 107,000,000 OFFACR - 468 DISCBD + 339,000,000 DELEMP + 16,200

MANF = 263 HWLMNX + 11.5 MPET2 - 9.05 MPE70 + 3,210 RRMI
+ 3,270,000,000 DELEMP + 383,000,000 MANACR - 2,090

HWLMNX = 0.0000308 RES - 0.000362 OFFICE + 0.000170 COMM - 7.39 HWYINT
- 0.246 DISCBD + 212,000 AUTO + 23.0

WHOLE = 3,948,000,000 WWEA + 2,000 HWLMNX + 1,670 RRMI + 6,000 HWYINT
- 104,000

HOTEL = 0.0524 MANF + 0.105 OFFICE + 1.69 MPE70 - 1.45 MPET2 - 233 DISCBD
- 2,420

HOSPTL = 191,000,000 NONHSE - 327 DISCBD + 10,000

CULTUR = 0.0023 RES - 14.4 DISCBD + 1,260
CHURCH = 0.00513 RES + 4,750
EDUC = 0.0405 RES + 28,200
REC = 0.649 RES + 979,000 MINCC - 745,000

F. COMPARISON OF MODEL FORMS

The analyses presented so far in this report have assumed a linear form for the model equations, e.g.

$$\text{RES} = b_1 \text{ OFFICE} + b_2 \text{ HWLMNX} + b_3 \text{ MPR70} + b_4 \text{ DUACRE} + c$$

This form assumes the independent effects of the model variables on RES are additive. An alternative assumption is that the effects are multiplicative, e.g.

$$\text{RES} = c * \text{OFFICE}^{b_1'} * \text{HWLMNX}^{b_2'} * \text{MPR70}^{b_3'} * \text{DUACRE}^{b_4'}$$

In order to test out this model form, the final path analysis was rerun using the natural logarithms of all data values in the Residential and Industrial/Office data files. Path diagrams for the multiplicative form are shown in Figures 6-9 and 6-10. Notice that the R^2 for the OFFICE equation in the Residential model is negative (Figure 6-9). This probably indicates that the sufficient condition for the two stage least squares is not satisfied (see Section III.D) or that some other singularity has been introduced into the problem when the equation is analyzed in multiplicative form.

The resultant R^2 values for the multiplicative form were in general much lower than those obtained using the additive form, however, in two cases, the multiplicative form did perform significantly better. In the Residential model, the R^2 for the EDUC equation rose significantly from 0.481 (additive) to 0.737 (multiplicative). The individual F-statistics and β weights for each independent variable rose as well. Therefore, the multiplicative form could be used for this particular equation. The final multiplicative equation is:

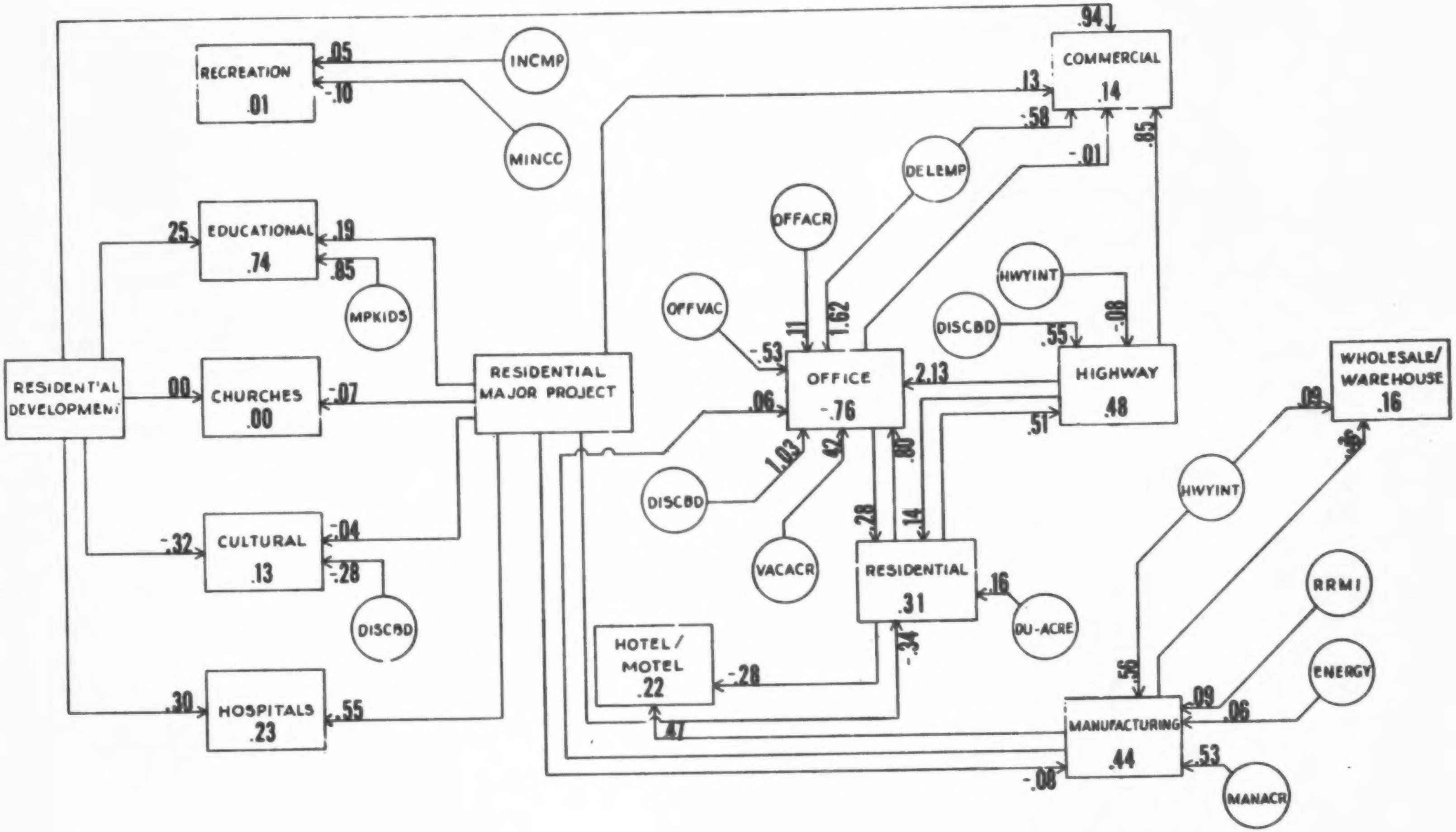

RESIDENTIAL DEVELOPMENT
RECREATION .01
EDUCATIONAL .74
CHURCHES .00
CULTURAL .13
HOSPITALS .23
INCMP
MINCC
MPKIDS
DISCBD
.94
.05
-.10
25
.19
.85
.00
-.07
-.32
-.04
-.28
.30
.55
RESIDENTIAL MAJOR PROJECT
COMMERCIAL .14
.13
.58
-.01
.85
DELEMP
OFFACR
OFFVAC
DISCBD
VACACR
HWYINT
DISCBD
OFFICE -.76
HIGHWAY .48
RESIDENTIAL .31
HOTEL / MOTEL .22
MANUFACTURING .44
WHOLESALE/ WAREHOUSE .16
.11
1.62
-.53
.06
1.03
.42
2.13
.08
.28
.14
.51
.55
-.08
.16
DU-ACRE
-.34
-.28
.47
-.08
HWYINT
RRMI
ENERGY
MANACR
.09
.36
.56
.09
.06
.53

Figure 6-9 Path Analysis for the Residential Model in Multiplicative Form

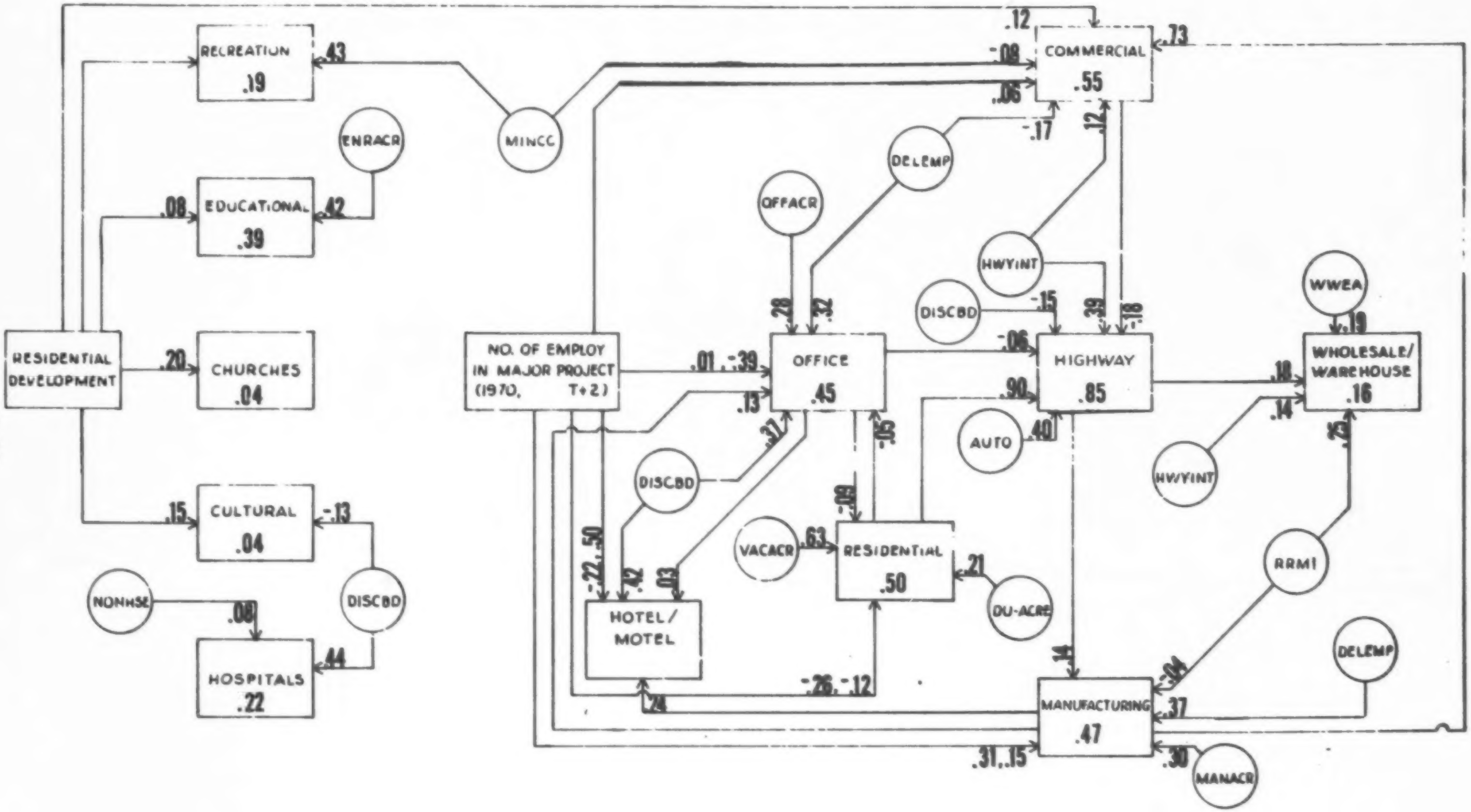

RESIDENTIAL DEVELOPMENT
RECREATION .19
EDUCATIONAL .39
CHURCHES .04
CULTURAL .04
HOSPITALS .22
ENRACR
MINCC
NONHSE
DISCBD
.12
.43
.08
.42
.20
.15
-.13
.08
.44
NO. OF EMPLOY IN MAJOR PROJECT (1970, T+2)
COMMERCIAL .55
OFFICE .45
HIGHWAY .85
WHOLESALE/ WAREHOUSE .16
RESIDENTIAL .50
HOTEL/ MOTEL
MANUFACTURING .47
OFFACR
DELEMP
HWYINT
DISCBD
AUTO
VACACR
DU-ACRE
RRMI
MANACR
WWEA
.73
-.08
.06
-.17
.12
.28
.32
-.15
.39
.18
-.06
.90
.40
.01, -.39
.13
.37
.05
-.09
.63
.21
-.22, .50
.42
.03
-.26, -.12
.24
.14
-.04
.37
.30
.31, .15
.19
.18
.14
.29

Figure 6-10 Path Analysis for the Industrial-Office Model in Multiplicative Form

$$EDUC = 0.234 * MPKIDS^{1.87} * RES^{0.157} * MPR70^{0.651} \text{ (English Units)}$$
$$EDUC = 0.00513 * MPKIDS^{1.87} * RES^{0.157} * MPR70^{0.651} \text{ (Metric Units)}$$

$R^2 = 0.737$
$R^2_a = 0.706$

In the second case, a higher R^2 was obtained for the HWLMNX equation in the Industrial/Office model by using the multiplicative form. The R^2 rose from 0.682 (additive) to 0.845 (multiplicative). Although this change also represents a significant improvement for a particular equation, in the current project, the existence of feedback loops between HWLMNX and other endogenous variables implies that all equations in the simultaneous block should be in one form or another. If some equations in the simultaneous block were in additive form and some in multiplicative form, then an enlarged set of independent variables (endogenous and instrumental) would be required in the model equation. If a given variable appears both in an additive and a multiplicative equation, then it would have to be represented by two variables in the simultaneous block (e.g., HWLMNX and log (HWLMNX)) because of the interconnections between equations. The additional independent variables would create a degrees of freedom problem in many of the model equations, preventing correct solution of the equations. The use of the multiplicative form exclusively in the RES, COMM, OFFICE, and MANF equations would significantly reduce the R^2 values in each of these equations. Therefore, it is recommended that the additive form be retained in this case.

G. STABILITY OF PATH COEFFICIENTS

The usefulness of any path analytic model is dependent on the stability of the path estimates. Im the current study, where only a small data sample (20) was available for analysis, the question of stability was especially important. There are at least two approaches to evaluating the stability of model estimates.

One approach involves confidence intervals [24]. Given a model equation:

$$Y = b_0 + \sum_{i=1}^{n} b_i X_i$$

confidence intervals can be specified in the form:

$$Y \pm tV(Y)^{\frac{1}{2}}$$

where t is the t-statistic for the regression of the model equation and V(Y) is the variance of Y. V(Y) can be expressed as:

$$V(U) = \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j \text{ Covariance } (b_i b_j)$$

The confidence interval is thus, a function of actual data values, $X_i X_j$. This approach was not used, as it does not, for our purposes, adequately consider individual observations. All of the necessary data for computing confidence intervals exists in Appendices C and D (published in the separate appendix to this volume) [42].

The second approach, the statistical technique chosen to investigate stability, is jackknifing. This procedure involved 20 separate analyses of each model equation using only 19 data samples. For each analysis, a different sample from the total set of 20 was excluded. This approach provided 20 separate estimates of each of the model coefficients, from which mean and standard deviations, and one-way frequency distributions were prepared. Two separate indices of coefficient stability were then developed for interpreting the data. The jackknifing analysis was applied only to the five simultaneous block equations in each model. Only these equations were analyzed because of the following reasons:

- the five equations include the majority of the total land use, and almost all the significant emission sources.

- the jackknife, as employed here, is a relatively expensive procedure. (e.g., twenty regressions are run for each equation).

In the jackknifing analysis, only the path regression coefficients (b) were used. Path coefficients (β) do not provide a valid basis for comparisons of systems operating on different data populations since the variance of the independent variables is included in the path coefficients and variance can change dramatically from sample to sample [26].

The means and standard deviations computed from the set of 20 values for each model coefficient are compared on Table 6-9 with the coefficient values from the final path analysis. The results show the jackknifing results and final path analysis path regression coefficients to be practically identical with differences between the two sets, in general, on the order of two to three percent. The standard deviation of the jackknifing coefficients is, on the average, 49 percent of the coefficient mean value, indicating appreciable variance in the coefficient values, but no extreme instability. One-way frequency distributions of the coefficient values are graphed in Appendix B for the ten model equations in the two simultaneous blocks. To more accurately quantify the stability of the path regression coefficients and to allow comparisons, between equations, two separate indices of stability were developed.

1. Extreme Value Index

The first stability index that was used is a statistical test for extreme values in a sample developed by Dixon [41]. The null hypothesis assumes the value farthest from the mean does not differ significantly from the range of values in the data set, i.e., is not an extreme value. Assumptions are made that the data are independent observations and normally distributed. The coefficient values that were tested in the current application, however, are a dependent set. The violation of this assumption does not invalidate our use of the test statistic as a guide for judging stability. The corresponding levels of significance probably need adjustment. The 20 values

TABLE 6-9

COMPARISON OF PATH REGRESSION COEFFICIENTS FROM THE FINAL PATH ANALYSIS AND THE JACKKNIFING STABILITY ANALYSIS FOR THE SIMULTANEOUS BLOCK OF EQUATIONS

		Path Regression Coefficients*		
		Final Path	Jackknifing	
Model Equation	Independent Variable	Analysis	Mean	Standard Deviation
Residential Model				
RES	OFFICE	-1.38	-1.39	0.896
	HWLMNX	168	167	16.4
	MPR70	-0.808	-0.819	0.105
	DUACRE	3930	3860	798
COMM	RES	0.0814	0.0814	0.015
	OFFICE	0.649	0.652	0.200
	HWLMNX	18.4	18.3	4.42
	MPR70	0.0976	0.097	0.017
	DELEMP	-692	-677	154
OFFICE	RES	-0.0319	-0.030	0.011
	HWLMNX	5.74	5.86	6.71
	MANF	0.0572	0.063	0.051
	VACACR	-0.0127	-0.013	0.013
	OFFVAC	-5.26	-4.91	7.06
	OFFACR	690	724	276
	DISCBD	-15.4	-14.7	13.1
	DELEMP	765	738	197
MANF	MANACR	1470	1750	1260
	HWYINT	316	331	103
	ENERGY	4100	4110	891
	MPR70	-0.176	-0.176	0.068
	RRMI	79.8	81.0	23.8
HWLMNX	RES	0.00247	0.003	0.001
	HWYINT	-3.31	-3.20	1.13
	DISCBD	-1.73	-1.72	0.165

* Shown to three significant digits, wherever possible.

TABLE 6-9 (CONTINUED)

COMPARISON OF PATH REGRESSION COEFFICIENTS FROM THE FINAL PATH ANALYSIS AND THE JACKKNIFING STABILITY ANALYSIS FOR THE SIMULTANEOUS BLOCK OF EQUATIONS

		Path Regression Coefficients*		
		Final Path	Jackknifing	
Model Equation	Independent Variable	Analysis	Mean	Standard Deviation
Industrial/Office Model				
RES	OFFICE	3.96	4.43	2.35
	MPET2	-0.859	0.860	0.249
	MPE70	0.392	0.397	0.090
	DUACRE	4560	4570	714
	VACACR	1.05	1.033	0.247
COMM	RES	0.0785	0.077	0.006
	MANF	0.413	0.400	0.057
	MPE70	0.0367	0.036	0.011
	HWYINT	39.1	39.7	19.5
	DELEMP	-2270	-2180	697
	MINCC	-290	-297	178
OFFICE	RES	0.011	0.011	0.002
	MANF	0.0178	0.024	0.021
	MPET2	0.0199	0.020	0.008
	MPE70	-0.0195	-0.019	0.005
	OFFACR	-285	-282	61.0
	DISCBD	-8.11	-7.68	3.35
	DELEMP	902	821	371
MANF	HWLMNX	4.55	4.26	18.9
	MPET2	0.124	0.121	0.037
	MPE70	-0.0974	-0.096	0.038
	RRMI	55.6	54.5	29.8
	DELEMP	8690	9190	4070
	MANACR	1020	1160	1050

* Shown to three significant digits, wherever possible.

TABLE 6-9 (CONTINUED)

COMPARISON OF PATH REGRESSION COEFFICIENTS FROM THE FINAL PATH ANALYSIS AND THE JACKKNIFING STABILITY ANALYSIS FOR THE SIMULTANEOUS BLOCK OF EQUATIONS

Model Equation	Independent Variable	Path Regression Coefficients* Final Path Analysis	Jackknifing Mean	Jackknifing Standard Deviation
Industrial/Office Model (Continued)				
HWLMNX	RES	0.00178	0.002	0.001
	OFFICE	-0.0209	-0.019	0.010
	COMM	0.00980	0.009	0.004
	HWYINT	-4.59	-4.55	0.531
	DISCBD	-0.246	-0.259	0.254
	AUTO	32.6	32.7	5.25

* Shown to three significant digits, wherever possible.

for each model coefficient were ordered from lowest to highest $X_1<X_2< . . . . <X_{19}<X_{20}$, and the following ratio was computed:

$$r_1 = \frac{X_3 - X_1}{X_{k-2} - X_1}, \; k = 20$$

The critical r_1 level tested for was 0.535, corresponding to the one percent significance level. If r_1 exceeds 0.535, then, given an independent sample, there is a 99 percent probability that the deviation of the lowest value X_1 from the range of values in the data set is significant and that X_1 is an extreme value. Again as a guide, if r_1 exceeds 0.535 in our samples, we will refer to the corresponding X_1 as an extreme value significant at the .01 level. A separate test on the highest data value X_{20} was also performed. A second ratio to test the highest data point was computed as follows:

$$r_{20} = \frac{X_k - X_{k-2}}{X_k - X_3}, \; k = 20$$

The values of r_1 and r_{20} for each model coefficient are shown on Table 6-10 where the ratio was found to be significant at the one percent level, the case number of the data sample excluded from the path analysis corresponding to the extreme value is indicated.

When an extreme value occurs in the data set of path regression coefficient values, it indicates that the exclusion of data from a particular case number is causing a significant change in the value of the coefficient. Normally one would test for a change in value when certain data were included, however, it was impractical to compute coefficients for all possible data combinations, and thus, detection of changes was accomplished by excluding each sample, one at a time. If exclusion of one sample from a least squares analysis causes a significant change in a coefficient's values, then that data sample, when included, is exerting a greater effect on the analysis results than other samples.

TABLE 6-10

EXTREME VALUE INDEX OF STABILITY FOR ALL PATH REGRESSION COEFFICIENTS IN THE SIMULTANEOUS BLOCKS OF EQUATIONS

Model Equation	Independent Variable	Extreme Value Ratio r_1	r_{20}	Case No.* Excluded if Extreme Value** X_1	X_{20}
Residential Model					
RES	OFFICE	0.233	0.370		
	HWLMNX	0.124	0.124		
	MPR70	0.400	0.576		29
	DUACRE	0.693	0.141	16	
COMM	RES	0.333	0.283		
	OFFICE	0.803	0.762	15	36
	HWLMNX	0.761	0.711	36	22
	MPR70	0.461	0.503		
	DELEMP	0.528	0.475		
OFFICE	RES	0.342	0.333		
	HWLMNX	0.282	0.526		
	MANF	0.100	0.876		37
	VACACR	0.454	0.354		
	OFFVAC	0.454	0.297		
	OFFACR	0.744	0.893	37	15
	DISCBD	0.144	0.559		37
	DELEMP	0.651	0.134	36	
MANF	MANACR	0.625	0.967	37	15
	HWYINT	0.616	0.734	40	16
	ENERGY	0.635	0.377	35	
	MPR70	0.739	0.247	19	
	RRMI	0.632	0.417	25	
HWLMNX	RES	0.000	0.000		
	HWYINT	0.726	0.890	10	14
	DISCBD	0.676	0.523	10	
Industrial/Office Model					
RES	OFFICE	0.263	0.373		
	MPET2	0.844	0.785	11	12
	MPE70	0.388	0.405		
	DUACRE	0.547	0.750	17	12
	VACACR	0.754	0.633	12	38

*See Table C-1 in Appendix C for definitions of case numbers [42]

**At the 1 percent significance level.

TABLE 6-10 (CONTINUED)

EXTREME VALUE INDEX OF STABILITY FOR ALL PATH REGRESSION COEFFICIENTS IN THE SIMULTANEOUS BLOCKS OF EQUATIONS

Model Equation	Independent Variable	Extreme Value Ratio		Case No.* Excluded if Extreme Value**	
		r_1	r_{20}	X_1	X_{20}
COMM	RES	0.297	0.301		
	MANF	0.329	0.352		
	MPE70	0.403	0.292		
	MWYINT	0.319	0.139		
	DELEMP	0.311	0.137		
	MINCC	0.296	0.215		
OFFICE	RES	0.197	0.496		
	MANF	0.368	0.294		
	MPET2	0.278	0.284		
	MPE70	0.530	0.189		
	OFFACR	0.434	0.335		
	DISCBD	0.410	0.609		11
	DELEMP	0.535	0.515		
MANF	HWLMNX	0.743	0.391	24	
	MPET2	0.443	0.213		
	MPE70	0.180	0.204		
	RRMI	0.437	0.519		
	DELEMP	0.566	0.897	17	24
	MANACR	0.182	0.813		24
HWLMNX	RES	1.00	1.00	17	38
	OFFICE	0.408	0.251		
	COMM	0.418	0.323		
	HWYINT	0.315	0.312		
	DISCBD	0.500	0.573		11
	AUTO	0.239	0.346		

*See Table C-1 in Appendix C for definitions of case numbers [42]

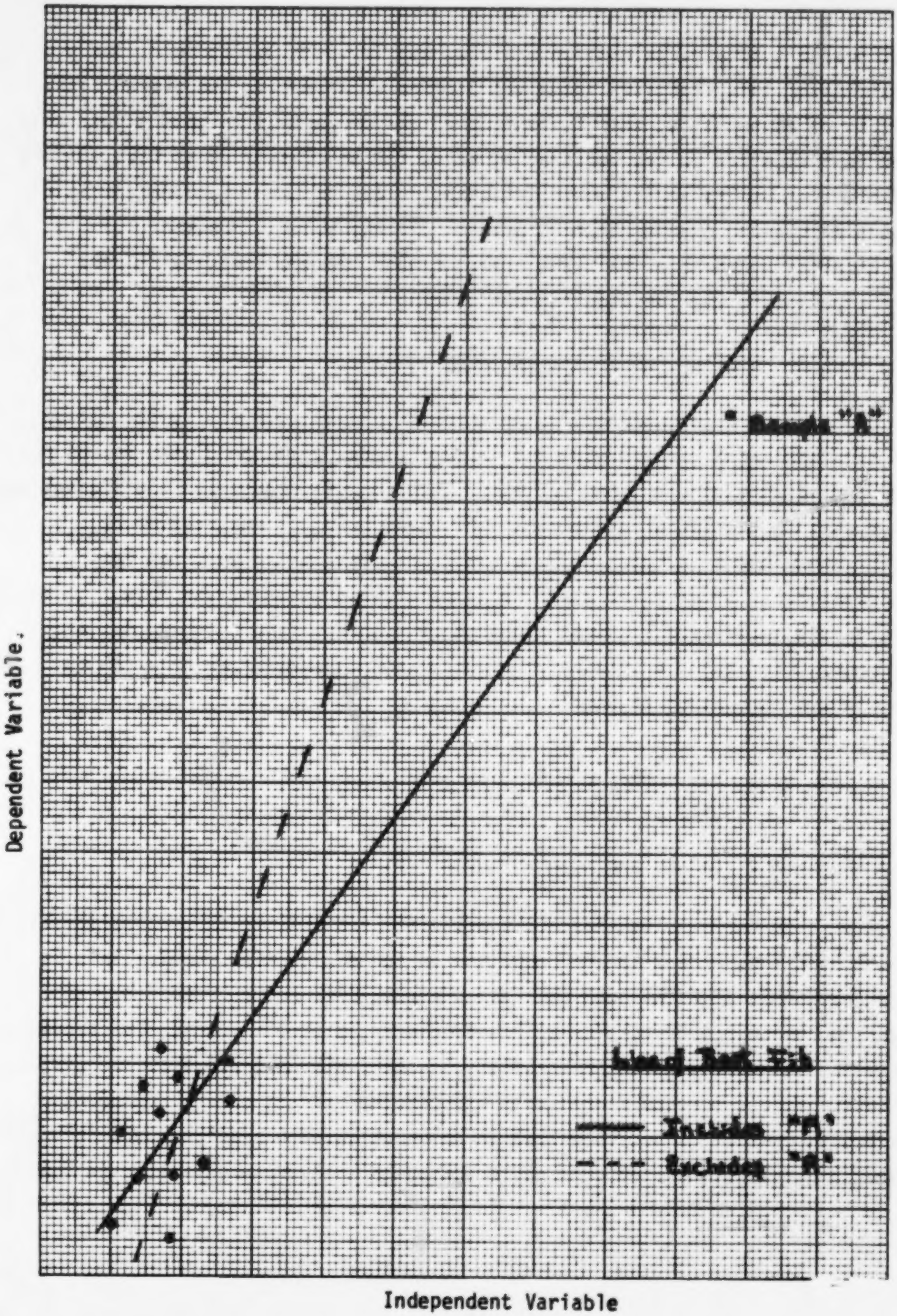
**At the 1 percent significance level.

An example of this situation is illustrated by a hypothetical simple regression in Figure 6-11. Here the data sample "A" is the only point in a particular portion of the variable range, and so its position allows it to exert a lot of leverage in the regression when it is included, causing a significant charge in the line of best fit, which would not occur with any of the other points. This does not mean that "A" is in error; it could just be the only data sample available in a particular portion of the variable range. To exclude it from the regression when it is a valid sample would, in fact, cause an error in the estimation of model coefficients. In the multiple regressions used in the path analysis, this situation is not so easily visualized. The anomalous sample may not be alone in a different range of values, but just far from the line of best fit determined using all other samples (dotted line in Figure 6-11).

In the path analysis, only 20 data samples were available for each regression and so it is not surprising with such a small sample size that instabilities occur in which particular samples exert significant changes in some of the model coefficients. When a particular case number reappears in the same model equation more than once, it is probably an indication that data for the dependent variable (and not the independent variables listed) is causing the extreme value. An example of this is the OFFICE equation in the Residential model where case number 37 appears three times (see Table 6-8). Thus, multiple occurrences of the same case number in a given equation is not as strong an indication of instability as the occurrence of the same number of different case numbers.

The existence of instability in a model coefficient may simply be an indication of the small sample size available for analysis, or it may indicate an exception to the general relationships developed in the final path analysis. In the first case, additional data in the analysis would remove the instability and probably give a mean value for the coefficients close to that obtained in the jackknifing analysis. In the second case, however, additional data would probably not correct the instability in the model. Since it was

FIGURE 6-11 HYPOTHETICAL LINEAR REGRESSION



not possible to ascertain which was the case for each coefficient, the validity of this stability index was limited to assessing the overall stability of an equation, not its individual coefficients. In this light, only the MANF equation in the Residential model can be classified as having appreciable instability in its model estimates. This is the only equation in Table 6-10 having greater than 50 percent extreme values (7 out of 10). It is interesting to note that this equation also has the lowest R^2 (0.407) of any equation in the two simultaneous blocks.

2. Coefficient Variation Index

A second stability index was developed to quantify the effect of the total variation in each model coefficient on the predicted values for the dependent variable. The variation in each path regression coefficient was expressed as the difference between the maximum (b_{max}) and minimum (b_{min}) values obtained in the jackknifing analysis. This difference was then multiplied by the mean value of the associated independent variable ($\bar{X}_i$) to obtain the total effect on the dependent variable. This quantity was then normalized by the mean value of the associated dependent variable ($\bar{Y}$) to obtain the percentage change:

$$P_i = \left| \frac{(b_{max} - b_{min})\bar{X}_i}{\bar{Y}} \right|$$

The computed P_i values are summarized in Table 6-11.

The results indicate substantial instability is present in all model equations except the RES equation in the Residential model. This is the only equation which does not have a p value exceeding 50 percent. The average percentage change for all coefficients is 90.3 percent. The above results deal solely with the stability of the individual regression coefficients. If one were concerned with the combined effect on the dependent variable, then it is possible that in an equation with many independent variables these individual effects could add up to an average variation in the

TABLE 6-11

STABILITY INDEX FOR ALL PATH REGRESSION COEFFICIENTS IN THE SIMULTANEOUS BLOCKS OF EQUATIONS BASED ON THE EFFECT OF VARIATION IN THE COEFFICIENTS ON THE DEPENDENT VARIABLE

Model Equation	Independent Variable	Percentage Change in Dependent Variable
Residential Model		
RES	OFFICE	22.5
	HWLMNX	27.4
	MPR70	36.9
	DUACRE	24.9
COMM	RES	27.9
	OFFICE	33.9
	HWLMNX	50.4
	MPR70	24.4
	DELEMP	15.9
OFFICE	RES	71.9
	HWLMNX	222.0
	MANF	74.8
	VACACR	134.0
	OFFVAC	71.4
	OFFACR	75.8
	DISCBD	343.0
	DELEMP	556.0
MANF	MANACR	111.0
	HWYINT	66.5
	ENERGY	493.0
	MPR70	119.0
	RRMI	55.3
HWLMNX	RES	20.9
	HWYINT	30.3
	DISCBD	72.8
Industrial/Office Model		
RES	OFFICE	19.9
	MPET2	49.1
	MPE70	30.9
	DUACRE	36.2
	VACACR	143.0

TABLE 6-11 (CONTINUED)

STABILITY INDEX FOR ALL PATH REGRESSION COEFFICIENTS IN THE SIMULTANEOUS BLOCKS OF EQUATIONS BASED ON THE EFFECT OF VARIATION IN THE COEFFICIENTS ON THE DEPENDENT VARIABLE

Model Equation	Independent Variable	Percentage Change in Dependent Variable
COMM	RES	18.5
	MANF	37.9
	MPE70	29.1
	HWYINT	167.0
	DELEMP	260.0
	MINCC	68.7
OFFICE	RES	42.3
	MANF	76.6
	MPET2	51.5
	MPE70	74.2
	OFFACR	29.8
	DISCBD	96.9
	DELEMP	112.0
MANF	HWLMNX	204.0
	MPET2	32.4
	MPE70	58.7
	RRMI	73.1
	DELEMP	146.0
	MANACR	93.6
HWLMNX	RES	79.1
	OFFICE	23.0
	COMM	56.5
	HWYINT	15.0
	DISCBD	41.9
	AUTO	19.5

dependent variable of several hundred percent. It is also possible that some of the effects could cancel each other reducing this variation.

Of the 21 different independent variables listed on Table 6-9, 17 of them (or 81%) are associated with a p value exceeding 50 percent. Thus, the instability problem is wide spread and not confined to a few variables in certain equations. The far extent of the problem suggests that the cause is probably the small sample size of 20 that was available for analysis. With a larger data set to work with, path regression coefficients should not change as much when a particular data sample is excluded, thus reducing the spread between b_{max} and b_{min} and the value of p.

H. CONCLUSIONS

A theoretical model of the Growth Effects of Major Land Use Projects has been developed. This model was keyed towards representing the total land use in a circular area of influence of 10,000 acres ten years after the construction of a Major Project in that area. The model was specified in two separate forms to represent induced land use growth associated with large Residential developments, and large Industrial or Office parks in the following 12 land use categories:

Residential
Commercial
Office
Manufacturing
Highways
Wholesale/Warehouse
Hotels/Motels
Hospitals
Cultural
Churches
Education
Recreation

The assumption of a single basic causal structure for induced development, and the use of cross-sectional data from diverse locations throughout the United States, allowed a static approach to the testing of the theoretical models, using path analysis.

Path analysis is a set of statistical techniques useful in testing theories and studying the logical consequences of various hypothesis involving causal relations. It is not capable of deducing or generating causal relations, only testing them. The causal analysis of induced land use development in the current study involved the use of two basic statistical techniques: two-stage least squares and stepwise ordinary least squares (multiple regression). The first technique was required to produce consistent estimates of the path coefficients in a system of simultaneous equations involving feedback loops or reciprocal causation in the models. The second technique was used to solve the remaining recursive portions of the models. The dependent variables in these regression analyses represented the total land use in the previously noted 12 categories. Both linear and non-linear forms were tested and the linear form was found to produce the best fit. Specific statistical criteria were developed to identify model paths that were insignificant or redundant, and these criteria were used to trim unneeded and undesirable paths from the models. A second complete path analysis was performed, and the trimming process repeated several times until the final path models were decided upon. The trimming process eliminated almost half of the paths in the models as originally specified.

The final models of land use development show that strong statistical relationships exist between the variables representing the 12 categories of total land use and the other model variables representing induced and non-induced land use growth processes. Only in the case of cultural land use did the path analysis reject the hypothesized causal relationships. Excluding this category, the R^2 statistic for the model equations in the simultaneous block ranged from 0.43 to 0.81 with an average value of 0.66 and the R^2 values for the model equations in the recursive block ranged from 0.12 to 0.86 with an average value of 0.41. These statistics can be interpreted as the amount of variance in the dependent variables (total land use) of the model equations that can be explained through the linear relationships in the final causal model. These results indicate a good verification of the hypothesized land use development model (and above the expectations of the authors).

There were several problems encountered in the path analysis, involving multicollinearity, suppressor variables, choice of instrumental variables, available degrees of freedom, and coefficient instability. The first two problems were eliminated through the approach used for theory trimming of the models. The last three problems were caused principally by a common element: the small number of data samples (20) available for analysis. In the model equations, as originally specified, there were sometimes as many independent variables as data samples. Since at least several degrees of freedom should be reserved for the error term in any multiple regression, some model paths had to be trimmed prior to, and in order to perform, the first path analysis. Thus, the limited data sample did preclude the testing of causal relationships in some instances. Also, an analysis of the stability of the model path coefficients revealed appreciable instability in the individual model coefficients when it was applied to different subsets of the original land use data set.

We note that this instability does not invalidate the strong causal relationships confirmed by the path analysis. These causal relationships will be translated into predictive equations in the subsequent task in this contract, the model calibration phase. This effort is reported in a forthcoming third volume of this report.

VII. REFERENCES

1. 40 *Federal Register* 40048 (October 20, 1975),
   40 *Federal Register* 41941 (September 9, 1975),
   40 *Federal Register* 25814 (June 19, 1975),
   40 *Federal Register* 23746 (June 2, 1975),
   40 *Federal Register* 18726 (April 29, 1975),
   40 *Federal Register* 16343 (May 8, 1974),
   40 *Federal Register* 9599 (April 18, 1973), and
   40 *Federal Register* 6279 (March 8, 1973).

2. U.S. Environmental Protection Agency, *Review of Federal Actions Impacting the Environment*, Washington, DC, EPA, 1975 (Manual TN2/3-1-75).

   Office of Federal Activities, U.S. Environmental Protection Agency, *Guidelines for Review of Environmental Impact Statements*; *Volume I, Highway Projects*, Washington DC, EPA, 1973, (Volume II on Airports and Volume III on Steam Channelization will be published shortly).

   39 Federal Register 16186, (May 7, 1974).

   Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, *Guidelines for Preparing Environmental Impact Statements*, Research Triangle Park NC, OAQPS, May 1975.

3. 40 *Federal Register* 28064 (July 3, 1975),
   39 *Federal Register* 45014 (December 30, 1974),
   39 *Federal Register* 25292 (July 9, 1974),
   39 *Federal Register* 7270 (February 24, 1974), and
   38 *Federal Register* 15834 (June 18, 1973).

4. 40 *Federal Register* 25504 (June 12, 1975),
   39 *Federal Register* 42510 (December 5, 1974),
   39 *Federal Register* 31000 (August 27, 1974),
   38 *Federal Register* 18986 (July 16, 1973), and
   37 *Federal Register* 23836 (November 9, 1972).

5. Lowry, Ira, *A Model of Metropolis*, Santa Monica, California, The RAND Corporation, 1964, (RM-4035-R6).

6. Forrester, Jay, *Urban Dynamics*, Cambridge, Massachusetts, The MIT Press, 1969.

7. Hill, Donald, "A Growth Allocation Model for the Boston Region" *Journal of the American Institute of Planners* (1965)[3], This is the EMPIRIC model.

8. Seidman, D.R., The Construction of an Urban Growth Model, Philadelphia, Delaware Valley Regional Planning Commission, 1969. This is the Penn-Jersey model.

9. Center for Real Estate and Urban Economics, Jobs, People, and Land, Bay Area Simulation Study, Berkeley, University of California, 1968, (Special Report No. 6). This is the BASS model.

10. Ohls, James, C., and Peter Hutchinson, "Models in Urban Development" pp. 165-200 in: Saul I. Gass and Roger L. Sisson (eds), A Guide to Models in Governmental Planning and Operations, Potomac, Maryland, Sauger Books, 1975.

11. Office of Air Quality Planning and Standards, Guidelines for Air Quality Maintenance Planning and Analysis; Volume 4, Land Use and Transportation Considerations, Research Triangle Park, NC, U.S. Environmental Protection Agency, 1974 (EPA-450/4-74-004).

12. Heise, David, R., Causal Analysis, Chapel Hill, NC, Unpublished manuscript, 1974.

13. Blalock, Hubert, M., Jr., Theory Construction, Englewood Cliffs, NJ, Prentice-Hall, 1969.

14. Meisel, William, S., and David C. Collins, "Repro-Modeling: An Approach to Efficient Model Utilization and Interpretation", IEEE Transactions on Systems, Man and Cybernetics (1973) 3:349-358.

15. D'Agostino, Ralph, "Path Analysis", Boston, Unpublished paper, 1975.

16. Wright, S., "The Method of Path Coefficients", Annals of Mathematical Statistics, 1934 (Vol. 5, pp. 161-215).

17. Wright, S., "The Interpretation of Multivariate Systems", Statistics and Mathematics in Biology, Edited by O. Kempthorne, T. Bancroft, J. Cowen, and J. Lush, Iowa State College Press, Ames, Iowa, 1954 (Vol. 11-33).

18. Wright, S., "Path Coefficients and Path Regressions Alternative or Complementary Concepts", Biometrics, 16, 1960 (pp. 189-202).

19. Wright, S., "The Treatment of Reciprocal Interaction, with or without Lag, in Path Analysis", Biometrics, 16, 1960 (pp. 423-445)

20. Van de Geer, J.P., Introduction to Multivariate Analysis for the Social Sciences, W.H. Freeman, San Francisco, California, 1971.

21. Kerlinger, F.N., and Pedhazer, E.J., Multiple Regression in Behavioral Research, Holt, Rinehart and Winston, New York, N.Y., 1973.

22. Heise, D.R., "Problems in Path Analysis and Causal Inference", Sociological Methodology, edited by E.F. Borgatta, Josey-Bass, San Francisco, California, 1969 (pp. 38-72).

23. Goldberger, A.S., Econometric Theory, Wiley, New York, NY, 1964.

24. Johnston, J., Econometric Methods, McGraw-Hill, New York, NY, 1963.

25. Snedecor, G.W., and Cochran, W.G., Statistical Methods (6th Edition), Iowa State College Press, Ames, Iowa, 1967.

26. Tukey, J.W., "Causation, Regression and Path Analysis", Statistics and Mathematics in Biology, Edited by O. Kempthorne, T.A. Bancroft, J.W. Gowen and J.L. Lush, Iowa State College Press, Ames, Iowa (pp. 35-66).

27. Turner, M.E., and Stevens, C.D., "The Regression Analysis of Causal Paths", Biometrics, 15, 1959 (pp. 236-258).

28. Heise, D.R., "Employing Nominal Variables, Induced Variables, and Block Variables in Path Analyses", Sociological Methods and Research, 1972 (pp. 147-173).

29. Land, K.C., "Principles of Path Analysis", Sociological Methodology, Edited by E.F. Borgatta, Jossey-Bass, San Francisco, California, 1969.

30. Boudon, R., "A Method of Linear Causal Analysis: Dependence Analysis", American Sociological Review 30, 1965 (pp. 365-373).

31. An Air Pollution Impact Methodology for Airports - Phase I, Argonne National Laboratory, Prepared for Office of Air and Water Programs, U.S., EPA, Durham, NC, January, 1973.

32. Urban Travel Patterns for Airports, Shopping Centers, and Industrial Plants, National Cooperative Highway Research Program Report, No. 24.

33. 1972 County and City Data Book, U.S. Bureau of the Census.

34. Survey of Buying Power, Sales Management Magazine, New York, NY (annual)

35. Wonnacott, Ronald, and Wonnacott, Thomas, Econometrics, John Wiley and Sons, 1970.

36. Metcalf and Eddy, Massachusetts Metropolitan District Commission Wastewater Study, 1973.

37. Lakshmanan and Hansen, "A Retail Market Potential Model", Journal of the American Institute of Planners, Vol. 31, May 1965.

38. Secondary Impacts of Transportation and Wastewater Improvements: Review and Bibliography, Prepared by the Environmental Impact Center for the Council on Environmental Quality and the U.S., EPA, 1975.

39. Infonet Times Series Processor, Information Network Division, Computer Sciences Corp., Los Angeles, California, December 1974.

40. Nie, N.H., C.H. Hull, J.G. Jenkins, K. Steinbrenner, and D.H. Bent, Statistical Package for the Social Sciences (second edition), McGraw-Hill, New York, 1975.

41. Dixon, W.J., "Analysis of Extreme Values", Annals of Mathematical Statistics, Vol. 21, pp. 488-506, 1950.

42. Benesh, Frank and Peter Guldberg, Growth Effects of Major Land Use Projects, Volume I: Appendices C and D, Prepared by Walden Research, Division of Abcor, for Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina. Available from Air Pollution Technical Information Center (MD#18), U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711 (APTIC Document #80998), May 1976.

43. Private Communication with Mr. Russell Hanson, Director of Fairfax County Industrial Authority, 1972.

APPENDIX A

BIBLIOGRAPHY OF LAND USE MODELS AND LAND USE DEVELOPMENT

Argonne National Laboratories, A General Methodology for the Planning and Analysis of Air Pollution Control Strategies based on Land Use, February, 1972.

Batty, Michael, "Recent Development in Land Use Modeling: A Review of British Research", Urban Studies, June, 1972.

Berry, Brian J.L. "The Retail Component of the Urban Model". Journal of the American Institute of Planners, Vol. XXXI, No. 2, May, 1965.

Central Regional Planning Commission, The Impact of New Housing on the Suburbs, Vol. I - The Apartments of Westborough, March 1972.

Chapin, F. Stuart Jr. "A Model for Simulating Residential Development," Journal of the American Institute of Planners, Vol. XXXI, No. 2, May, 1975.

Chapin, F. Stuart Jr. and Shirley F. Weiss, "Factors Influencing Land Development", Institute for Research in Social Science, Univer. of North Carolina, Chapel Hill, August, 1962.

Chorley, Richard J. and Peter Haggett, Socio-Economic Models in Geography, 1967.

Dickey, J.W. and P.W. Shuldiner, "Theory and Methodology Regional Planning Case Study 2: Maximum Traffic Generation for Planned Shopping Center" from paper presented at Annual Meeting of Highway Research Board, January, 1966.

Goldner, William, "The Lowry Model Heritage," Journal of the American Institute of Planners, Vo. XXXVI: No. 2, March, 1971.

Hill, Donald M. "A Growth Allocation Model for the Boston Region", Journal of the American Institute of Planners, Vol. 31:2, May, 1965.

Isard, Walter, "Ecologic-economic Analysis for Regional Development," 1972.

Highway Research Record, No. 126, "Land Use Forecasting Concepts, 7 Reports". 1966.

Lakshmanan, T.R. and Walter G. Hansen, "A Retail Market Potential Model", Journal of the American Institute of Planners, Vol. XXXI, May, 1965.

Lathrop, George T. and John R. Hamburg, "An Opportunity-Accessibility Model Allocating Regional Growth". Journal of American Institute of Planners Vol. XXXI No. 2, May, 1965.

Lee, Donald, Models and Techniques for Urban Planning, Cornell University, New York, September, 1968.

Leontif, Wassily, "Environmental Repercussions and the Economic Structure: An Input-Output Approach, "The Review of Economics and Statistics, Vol. 611, No. 3, August, 1970.

Lowry, C.S. "A Short Course in Model Design". Journal of the American Institute of Planners, Vol. XXXI, No. 2, May, 1965.

Metcalf & Eddy, Inc., MDC Wastewater Study, 1973.

Pack, Janet Rothenberg, "The Use of Urban Models: Report on a Survey of Planning Organizations". Journal of American Institute of Planners, May, 1975.

Roberts, John L. and Croke, Edward J. "A Critical Review of the Effect of Air Pollution Control Regulations on Land Use Planning." Journal of Air Pollution Control Association. May, 1975, Vol. 25, No. 5.

Steinitz, Carl and Peter Rogers, A System Analysis Model of Urbanization and Change, 1970.

Seidman, David R., "The Construction of an Urban Growth Model; A description of the Issues Confronted in the Development of the Activities Allocation Model", Delaware Valley Regional Planning Commission, 1970.

U.S. Environmental Protection Agency, 45014-74-004, Guidelines for Air Quality Maintenance Planning and Analysis, Volume 4: Land Use and Transportation and Wastewater Investments: Review and Bibliography, January, 1975.

U.S. Environmental Protection Agency, 68-61-0788. A Guide to Models in Governmental Planning and Operations, August, 1974.

U.S. Naval Civil Engineering Laboratory, The Impact of Large Installations on Nearby Areas - August, 1965.

APPENDIX B

PLOTS OF REGRESSION COEFFICIENTS FOR THE TWENTY JACKKNIFING RUNS

It is important to note that the scale for each graph is different, so that comparisons should not be made visually between the spread of points in different graphs. The scale of each graph is determined by the minimum and maximum values in the data set which are always plotted along the right- and left-hand edges of the graph, respectively. Data points are plotted useng the "*" character, except where coincident values occur, in which case a number indicating the multiplicity is plotted, e.g., "3" for three coincident points. Because of the differences in scale, data points whose values are grouped quite closely may appear to be spread out, and vice-versa. Thus, in order to quantify more accurately the stability of the path regression coefficients and to allow comparisons, the two separate indices of stability discussed in the text (page 6-66) were developed.

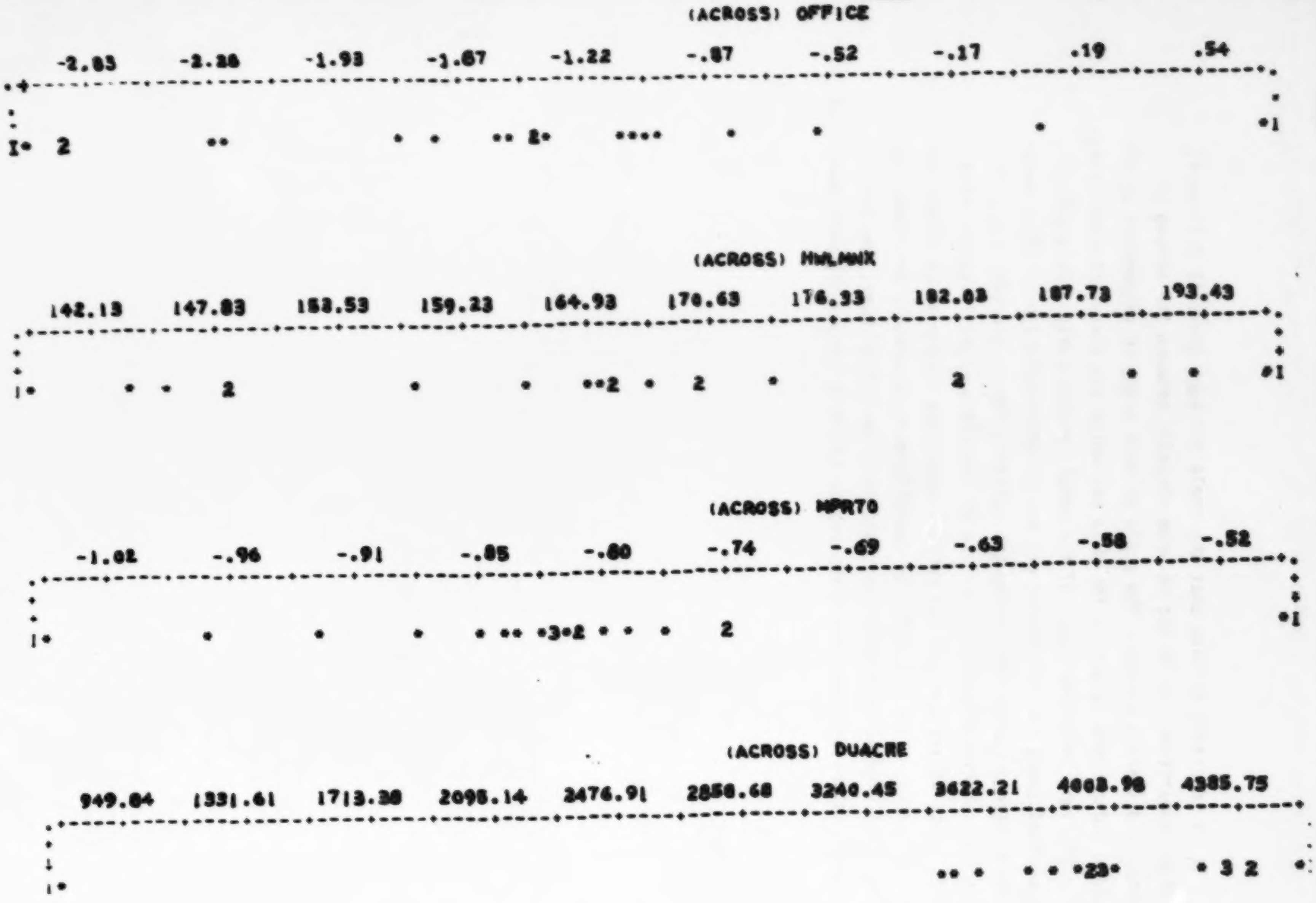


FIGURE B-1 PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE RES EQUATION IN THE RESIDENTIAL MODEL

"Please see the text for an explanation of the scales of the graphs"

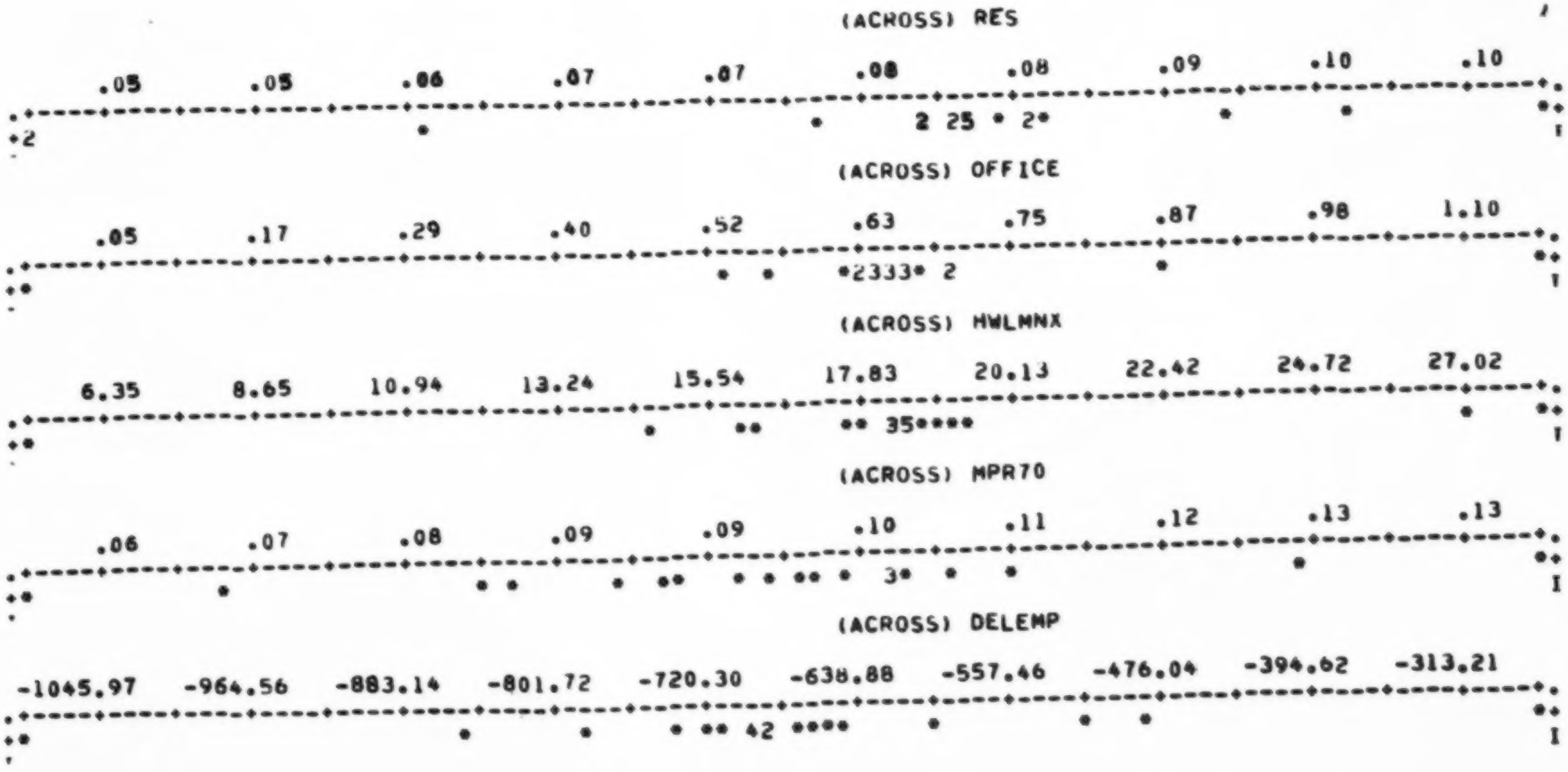


FIGURE B-2 PLOT OF-REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE COMM EQUATION IN THE RESIDENTIAL MODEL

"Please see the text for an explanation of the scales of the graphs"

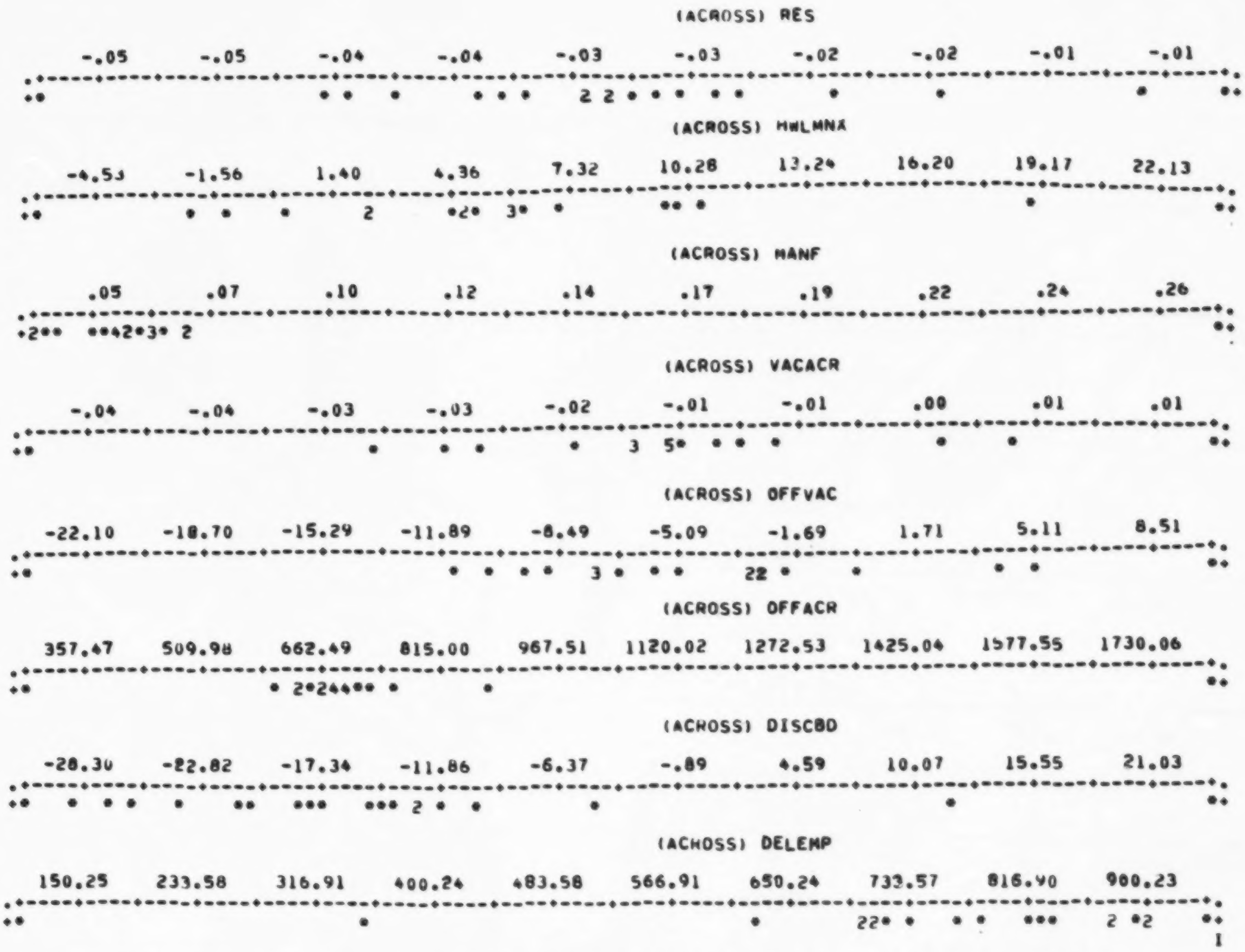


FIGURE B-3 PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE OFFICE EQUATION IN THE RESIDENTIAL MODEL

"Please see the text for an explanation of the scales of the graphs"

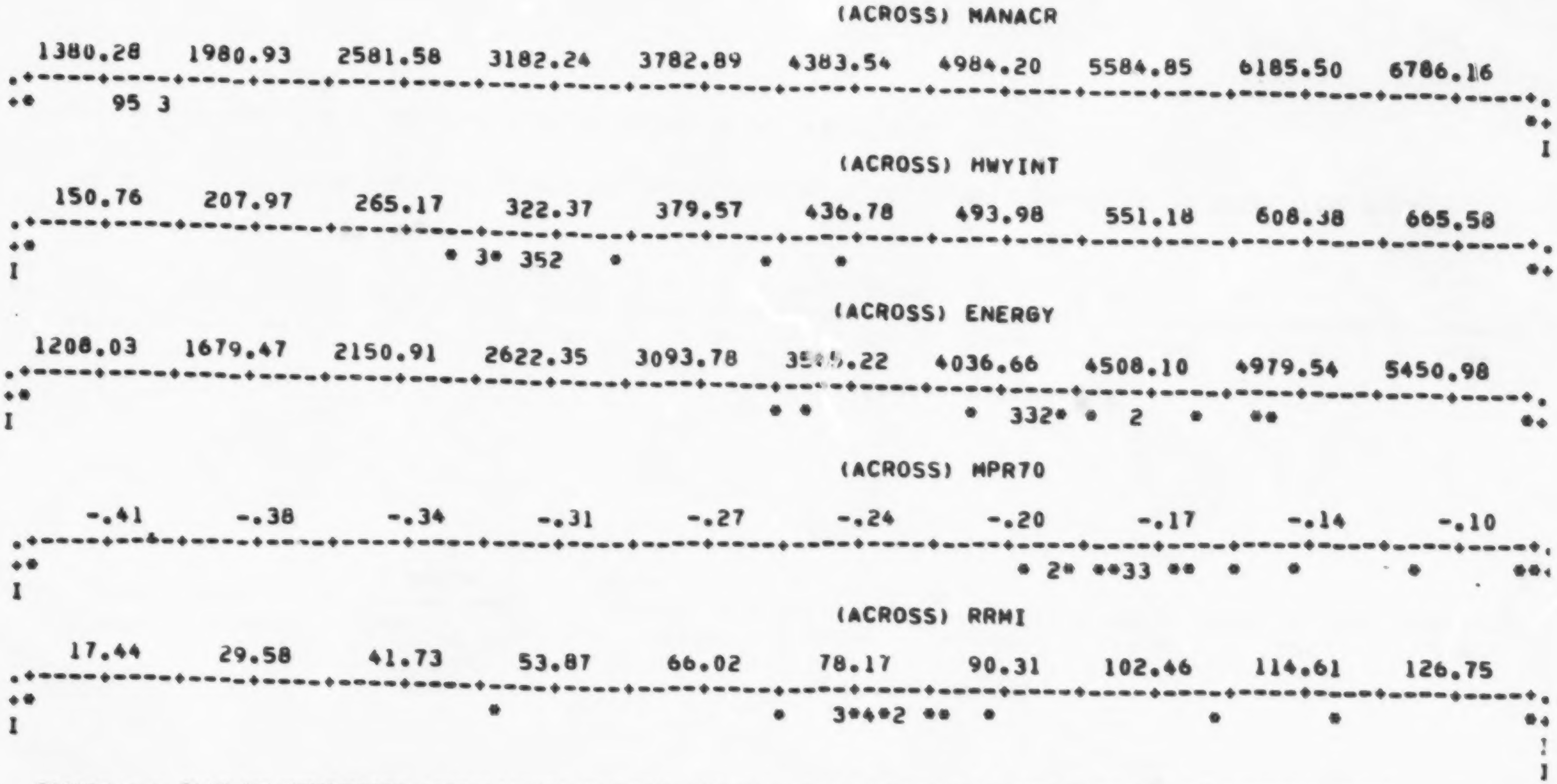


FIGURE B-4 PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE MANF EQUATION IN THE RESIDENTIAL MODEL

"Please see the text for an explanation of the scales of the graphs"

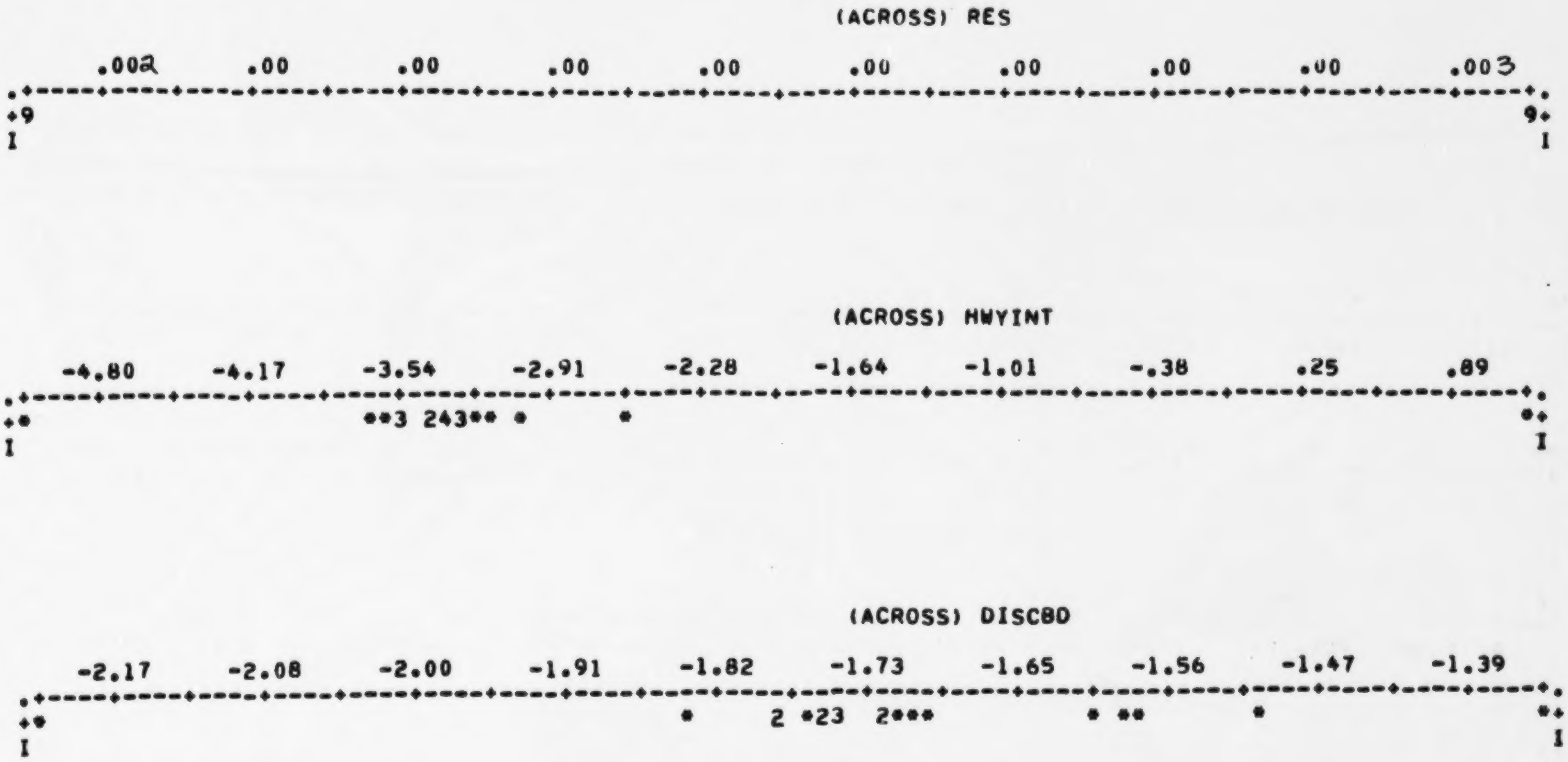


FIGURE B-5 PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE HWLMNX EQUATION IN THE RESIDENTIAL MODEL

"Please see the text for an explanation of the scales of the graphs"

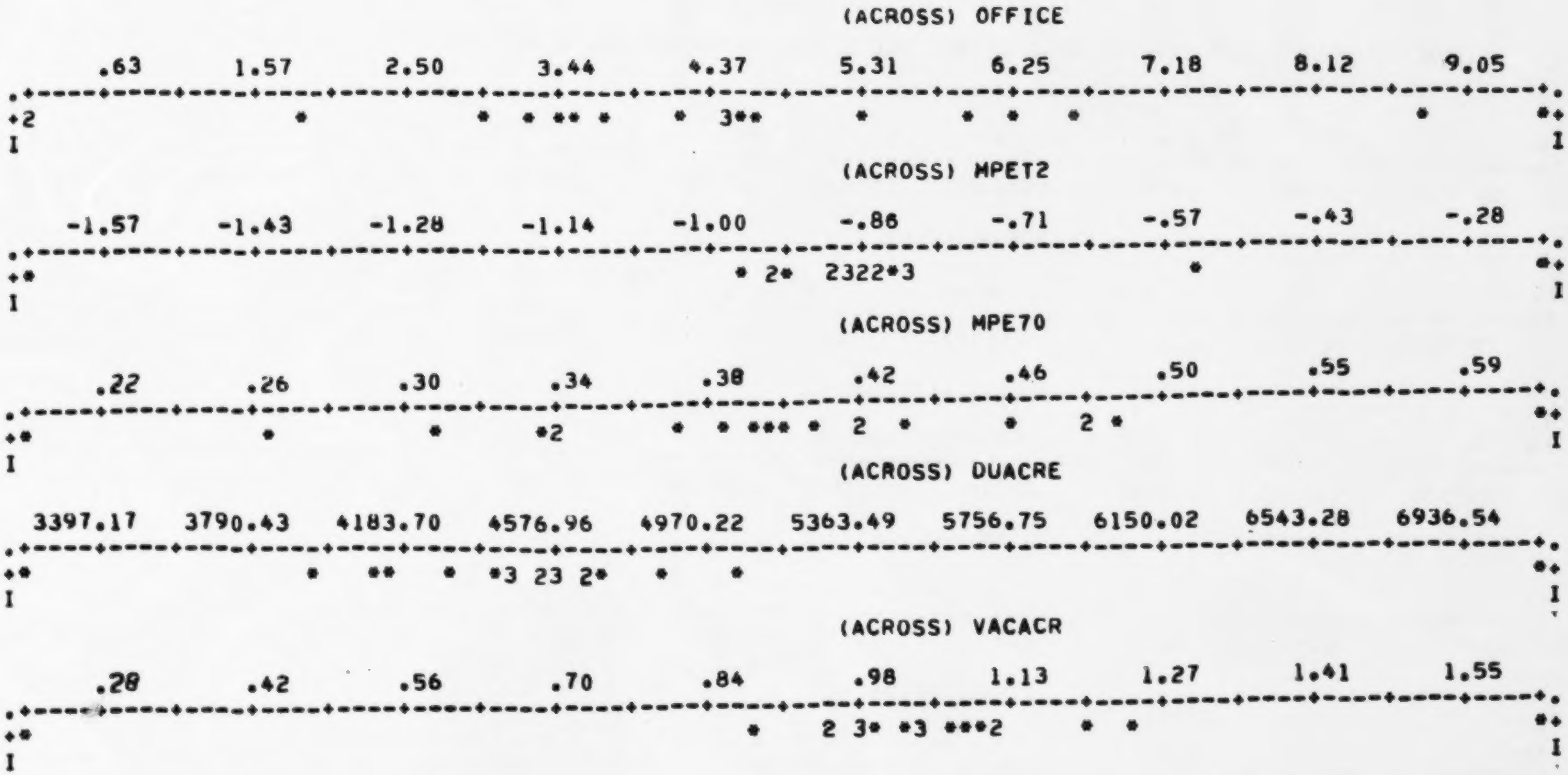


FIGURE B-6 PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE RES EQUATION IN THE INDUSTRIAL/OFFICE MODEL

"Please see the text for an explanation of the scales of the graphs"

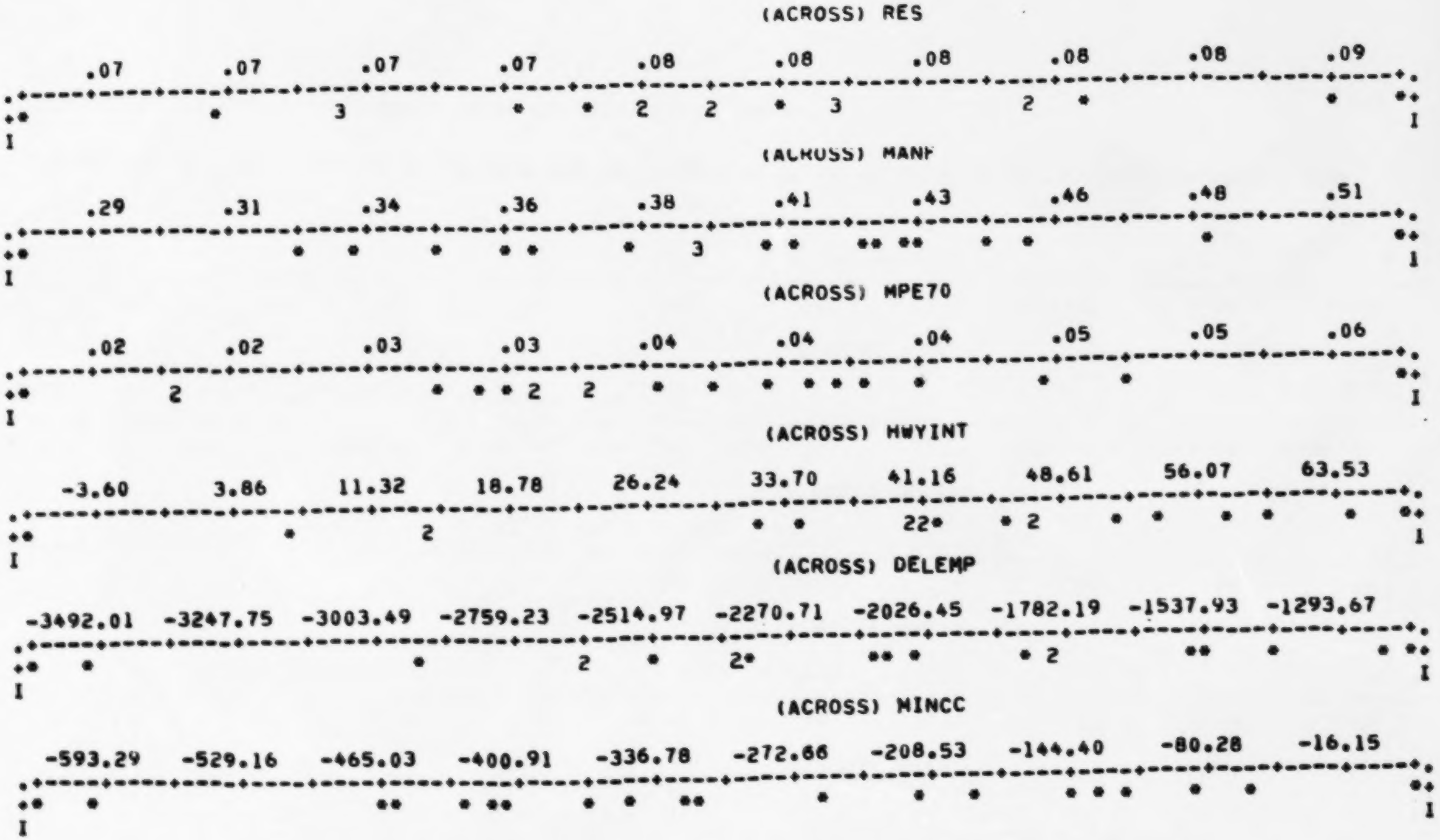


FIGURE B-7 PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE COMM EQUATION IN THE INDUSTRIAL/OFFICE MODEL

"Please see the text for an explanation of the scales of the graphs"

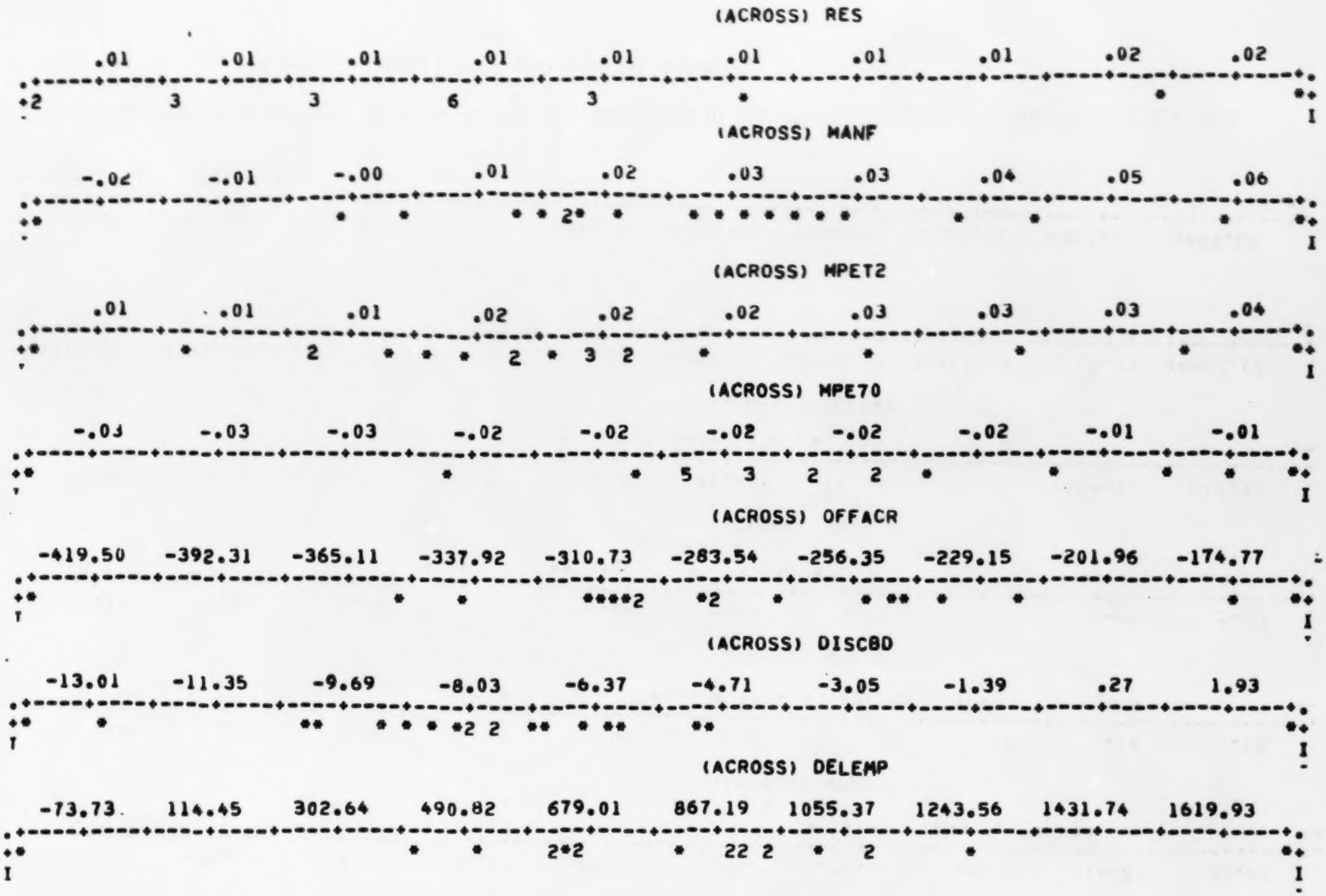


FIGURE B-8 PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE OFFICE EQUATION IN THE INDUSTRIAL/OFFICE MODEL

"Please see the text for an explanation of the scales of the graphs"

```
                                                       (ACROSS) HWLMNX

   -62.03    -52.17    -42.32    -32.47    -22.61    -12.76     -2.90     6.95      16.81     26.66
.+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+.
**                                                            *  * *    2*3 *2 **   ** **            **

                                                       (ACROSS) MPET2

     .02       .04       .06       .08       .09       .11       .13       .14       .16       .18
.+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+.
**                                *   *           *    * ** * *  22  **   * *           *   *        **

                                                       (ACROSS) MPE70

    -.16      -.14      -.13      -.11      -.10      -.08      -.07      -.06      -.04      -.03
.+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+.
** *            *    *          *  *      **  *2 *   *   ** *  *                    *          *     **

                                                       (ACROSS) RRMI

    -3.98     9.79    23.55    37.32    51.09    64.85    78.62    92.38   106.15   119.92
.+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+.
**  *                       *      *       **   *22 ****  * *  *          *                          **

                                                       (ACROSS) DELEMP

  5869.31   7975.29  10081.27  12187.25  14293.23  16399.21  18505.19  20611.16  22717.14  24823.12
.+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+.
**          ** 22 *4*32 *                                                                            **

                                                       (ACROSS) MANACR

   549.10   1051.37   1553.65   2055.93   2558.21   3060.49   3562.77   4065.05   4567.32   5069.60
.+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+.
** * * **2*   42*  * * *         *                                                                   **
```

FIGURE B-9. PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE MANF EQUATION IN THE INDUSTRIAL/OFFICE MODEL

"Please see the text for an explanation of the scales of the graphs"

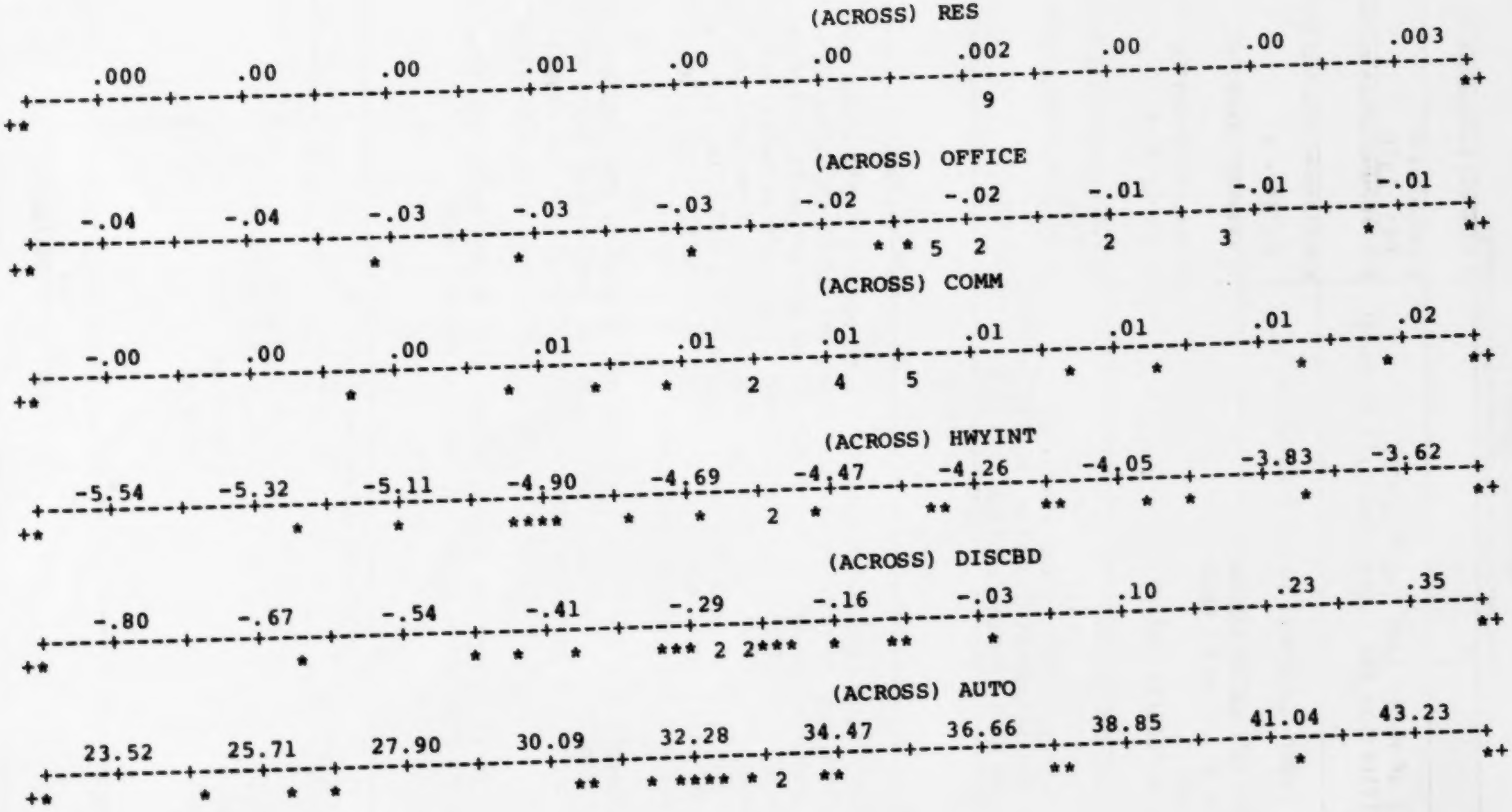


FIGURE B-10 PLOT OF REGRESSION COEFFICIENTS FOR THE JACKKNIFING OF THE HWLMNX EQUATION IN THE INDUSTRIAL OFFICE MODEL

"Please see the text for an explanation of the scales of the graphs"